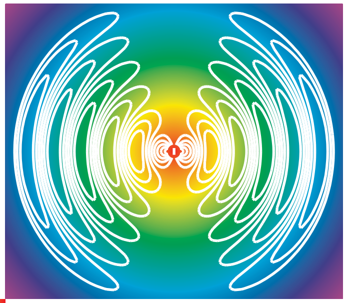
Electromagnetic Waves and Antennas

Sophocles J. Orfanidis



- Maxwell's Equations
- Uniform Plane Waves
- Pulse Propagation in Dispersive Media
- Slow, Fast, Negative Group Velocity
- Dispersion Compensation
- Pulse Compression and Chirp Radar
- Propagation in Birefringent Media
- Reflection and Transmission
- Multilayer Dielectric Structures
- Oblique Incidence
- Negative-Index Media
- Perfect Lenses
- Multilayer Film Applications
- Surface Plasmon Resonance
- Omnidirectional Dielectric Mirrors
- Giant Birefringent Optics
- Waveguides
- Transmission Lines
- Coupled Lines
- Impedance Matching
- S-Parameters
- Microwave Amplifier Design
- Radiation Fields
- Transmitting and Receiving Antennas
- Linear and Loop Antennas
- Radiation from Apertures
- Fresnel Diffraction
- Plane Wave Spectrum Representation
- Fourier Optics
- Aperture Antennas
- Horn and Patch Antennas
- Parabolic Antennas
- Antenna Arrays
- Array Design Methods
- Currents on Linear Antennas
- Hallen and Pocklington Equations
- Coupled Antennas

- MATLAB Functions

www.ece.rutgers.edu/~orfanidi/ewa

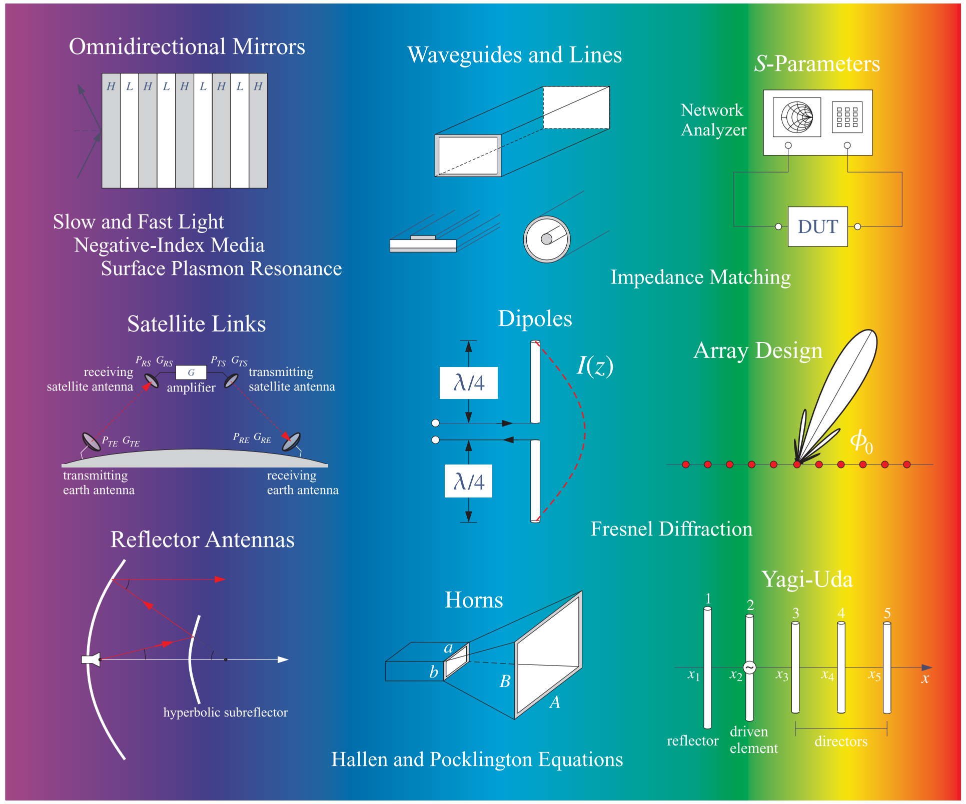
ISBN 978-0-9793713-2-5
90000
9 780979 371325

Orfanidis

Electromagnetic Waves and Antennas

Electromagnetic Waves and Antennas

Sophocles J. Orfanidis



Electromagnetic Waves and Antennas

Electromagnetic Waves and Antennas

Sophocles J. Orfanidis

Rutgers University

To Monica and John

Copyright © 1999–2008 by Sophocles J. Orfanidis

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.

MATLAB® is a registered trademark of The MathWorks, Inc.

Web page: `www.ece.rutgers.edu/~orfanidi/ewa`

Contents

Preface xiii

1 Maxwell’s Equations 1

1.1 Maxwell’s Equations, 1
1.2 Lorentz Force, 2
1.3 Constitutive Relations, 3
1.4 Boundary Conditions, 7
1.5 Currents, Fluxes, and Conservation Laws, 9
1.6 Charge Conservation, 10
1.7 Energy Flux and Energy Conservation, 11
1.8 Harmonic Time Dependence, 13
1.9 Simple Models of Dielectrics, Conductors, and Plasmas, 16
1.10 Kramers-Kronig Dispersion Relations, 26
1.11 Group Velocity, Energy Velocity, 29
1.12 Problems, 31

2 Uniform Plane Waves 36

2.1 Uniform Plane Waves in Lossless Media, 36
2.2 Monochromatic Waves, 42
2.3 Energy Density and Flux, 45
2.4 Wave Impedance, 46
2.5 Polarization, 46
2.6 Uniform Plane Waves in Lossy Media, 52
2.7 Propagation in Weakly Lossy Dielectrics, 58
2.8 Propagation in Good Conductors, 59
2.9 Propagation in Oblique Directions, 61
2.10 Complex or Inhomogeneous Waves, 63
2.11 Doppler Effect, 66
2.12 Propagation in Negative-Index Media, 70
2.13 Problems, 73

3 Pulse Propagation in Dispersive Media 80

3.1 Propagation Filter, 80
3.2 Front Velocity and Causality, 82
3.3 Exact Impulse Response Examples, 85
3.4 Transient and Steady-State Behavior, 88

3.5 Pulse Propagation and Group Velocity, 92
3.6 Group Velocity Dispersion and Pulse Spreading, 95
3.7 Propagation and Chirping, 100
3.8 Dispersion Compensation, 101
3.9 Slow, Fast, and Negative Group Velocities, 103
3.10 Chirp Radar and Pulse Compression, 110
3.11 Further Reading, 120
3.12 Problems, 120

4 Propagation in Birefringent Media 129

4.1 Linear and Circular Birefringence, 129
4.2 Uniaxial and Biaxial Media, 130
4.3 Chiral Media, 132
4.4 Gyrotropic Media, 135
4.5 Linear and Circular Dichroism, 136
4.6 Oblique Propagation in Birefringent Media, 137
4.7 Problems, 144

5 Reflection and Transmission 150

5.1 Propagation Matrices, 150
5.2 Matching Matrices, 154
5.3 Reflected and Transmitted Power, 157
5.4 Single Dielectric Slab, 160
5.5 Reflectionless Slab, 163
5.6 Time-Domain Reflection Response, 171
5.7 Two Dielectric Slabs, 173
5.8 Reflection by a Moving Boundary, 175
5.9 Problems, 178

6 Multilayer Structures 183

6.1 Multiple Dielectric Slabs, 183
6.2 Antireflection Coatings, 185
6.3 Dielectric Mirrors, 190
6.4 Propagation Bandgaps, 201
6.5 Narrow-Band Transmission Filters, 201
6.6 Equal Travel-Time Multilayer Structures, 206
6.7 Applications of Layered Structures, 220
6.8 Chebyshev Design of Reflectionless Multilayers, 223
6.9 Problems, 231

7 Oblique Incidence 238

7.1 Oblique Incidence and Snel’s Laws, 238
7.2 Transverse Impedance, 240
7.3 Propagation and Matching of Transverse Fields, 243
7.4 Fresnel Reflection Coefficients, 245
7.5 Maximum Angle and Critical Angle, 247
7.6 Brewster Angle, 256

7.7 Complex Waves, 258
7.8 Total Internal Reflection, 261
7.9 Oblique Incidence on a Lossy Medium, 262
7.10 Zenneck Surface Wave, 267
7.11 Surface Plasmons, 269
7.12 Oblique Reflection from a Moving Boundary, 272
7.13 Geometrical Optics, 276
7.14 Fermat's Principle, 279
7.15 Ray Tracing, 281
7.16 Snel's Law in Negative-Index Media, 292
7.17 Problems, 295

8 Multilayer Film Applications 300

8.1 Multilayer Dielectric Structures at Oblique Incidence, 300
8.2 Lossy Multilayer Structures, 302
8.3 Single Dielectric Slab, 304
8.4 Frustrated Total Internal Reflection, 306
8.5 Surface Plasmon Resonance, 310
8.6 Perfect Lens in Negative-Index Media, 319
8.7 Antireflection Coatings at Oblique Incidence, 327
8.8 Omnidirectional Dielectric Mirrors, 330
8.9 Polarizing Beam Splitters, 341
8.10 Reflection and Refraction in Birefringent Media, 343
8.11 Brewster and Critical Angles in Birefringent Media, 347
8.12 Multilayer Birefringent Structures, 350
8.13 Giant Birefringent Optics, 352
8.14 Problems, 357

9 Waveguides 359

9.1 Longitudinal-Transverse Decompositions, 360
9.2 Power Transfer and Attenuation, 365
9.3 TEM, TE, and TM modes, 367
9.4 Rectangular Waveguides, 370
9.5 Higher TE and TM modes, 372
9.6 Operating Bandwidth, 374
9.7 Power Transfer, Energy Density, and Group Velocity, 375
9.8 Power Attenuation, 377
9.9 Reflection Model of Waveguide Propagation, 380
9.10 Resonant Cavities, 382
9.11 Dielectric Slab Waveguides, 384
9.12 Problems, 392

10 Transmission Lines 394

10.1 General Properties of TEM Transmission Lines, 394
10.2 Parallel Plate Lines, 400
10.3 Microstrip Lines, 401
10.4 Coaxial Lines, 405
10.5 Two-Wire Lines, 410

10.6 Distributed Circuit Model of a Transmission Line, 412
10.7 Wave Impedance and Reflection Response, 414
10.8 Two-Port Equivalent Circuit, 416
10.9 Terminated Transmission Lines, 417
10.10 Power Transfer from Generator to Load, 420
10.11 Open- and Short-Circuited Transmission Lines, 422
10.12 Standing Wave Ratio, 425
10.13 Determining an Unknown Load Impedance, 427
10.14 Smith Chart, 431
10.15 Time-Domain Response of Transmission Lines, 435
10.16 Problems, 442

11 Coupled Lines 453

11.1 Coupled Transmission Lines, 453
11.2 Crosstalk Between Lines, 459
11.3 Weakly Coupled Lines with Arbitrary Terminations, 462
11.4 Coupled-Mode Theory, 464
11.5 Fiber Bragg Gratings, 466
11.6 Diffuse Reflection and Transmission, 469
11.7 Problems, 471

12 Impedance Matching 473

12.1 Conjugate and Reflectionless Matching, 473
12.2 Multisection Transmission Lines, 475
12.3 Quarter-Wavelength Chebyshev Transformers, 476
12.4 Two-Section Dual-Band Chebyshev Transformers, 482
12.5 Quarter-Wavelength Transformer With Series Section, 488
12.6 Quarter-Wavelength Transformer With Shunt Stub, 491
12.7 Two-Section Series Impedance Transformer, 493
12.8 Single Stub Matching, 498
12.9 Balanced Stubs, 502
12.10 Double and Triple Stub Matching, 504
12.11 L-Section Lumped Reactive Matching Networks, 506
12.12 Pi-Section Lumped Reactive Matching Networks, 509
12.13 Reversed Matching Networks, 516
12.14 Problems, 518

13 S-Parameters 522

13.1 Scattering Parameters, 522
13.2 Power Flow, 526
13.3 Parameter Conversions, 527
13.4 Input and Output Reflection Coefficients, 528
13.5 Stability Circles, 530
13.6 Power Gains, 536
13.7 Generalized S-Parameters and Power Waves, 542
13.8 Simultaneous Conjugate Matching, 546
13.9 Power Gain Circles, 551
13.10 Unilateral Gain Circles, 552

13.11 Operating and Available Power Gain Circles, 554
13.12 Noise Figure Circles, 560
13.13 Problems, 565

14 Radiation Fields 568

14.1 Currents and Charges as Sources of Fields, 568
14.2 Retarded Potentials, 570
14.3 Harmonic Time Dependence, 573
14.4 Fields of a Linear Wire Antenna, 575
14.5 Fields of Electric and Magnetic Dipoles, 577
14.6 Ewald-Oseen Extinction Theorem, 582
14.7 Radiation Fields, 587
14.8 Radial Coordinates, 590
14.9 Radiation Field Approximation, 592
14.10 Computing the Radiation Fields, 593
14.11 Problems, 595

15 Transmitting and Receiving Antennas 598

15.1 Energy Flux and Radiation Intensity, 598
15.2 Directivity, Gain, and Beamwidth, 599
15.3 Effective Area, 604
15.4 Antenna Equivalent Circuits, 608
15.5 Effective Length, 610
15.6 Communicating Antennas, 612
15.7 Antenna Noise Temperature, 614
15.8 System Noise Temperature, 618
15.9 Data Rate Limits, 624
15.10 Satellite Links, 626
15.11 Radar Equation, 629
15.12 Problems, 631

16 Linear and Loop Antennas 634

16.1 Linear Antennas, 634
16.2 Hertzian Dipole, 636
16.3 Standing-Wave Antennas, 638
16.4 Half-Wave Dipole, 642
16.5 Monopole Antennas, 643
16.6 Traveling-Wave Antennas, 645
16.7 Vee and Rhombic Antennas, 647
16.8 Loop Antennas, 650
16.9 Circular Loops, 652
16.10 Square Loops, 654
16.11 Dipole and Quadrupole Radiation, 655
16.12 Problems, 657

17 Radiation from Apertures 658

17.1 Field Equivalence Principle, 658
17.2 Magnetic Currents and Duality, 660
17.3 Radiation Fields from Magnetic Currents, 662
17.4 Radiation Fields from Apertures, 663
17.5 Huygens Source, 666
17.6 Directivity and Effective Area of Apertures, 668
17.7 Uniform Apertures, 670
17.8 Rectangular Apertures, 670
17.9 Circular Apertures, 672
17.10 Vector Diffraction Theory, 675
17.11 Extinction Theorem, 679
17.12 Vector Diffraction for Apertures, 681
17.13 Fresnel Diffraction, 682
17.14 Knife-Edge Diffraction, 686
17.15 Geometrical Theory of Diffraction, 694
17.16 Rayleigh-Sommerfeld Diffraction Theory, 700
17.17 Plane-Wave Spectrum Representation, 703
17.18 Fresnel Diffraction and Fourier Optics, 708
17.19 Problems, 719

18 Aperture Antennas 723

18.1 Open-Ended Waveguides, 723
18.2 Horn Antennas, 727
18.3 Horn Radiation Fields, 729
18.4 Horn Directivity, 734
18.5 Horn Design, 737
18.6 Microstrip Antennas, 740
18.7 Parabolic Reflector Antennas, 746
18.8 Gain and Beamwidth of Reflector Antennas, 748
18.9 Aperture-Field and Current-Distribution Methods, 751
18.10 Radiation Patterns of Reflector Antennas, 754
18.11 Dual-Reflector Antennas, 763
18.12 Lens Antennas, 766

19 Antenna Arrays 768

19.1 Antenna Arrays, 768
19.2 Translational Phase Shift, 768
19.3 Array Pattern Multiplication, 770
19.4 One-Dimensional Arrays, 780
19.5 Visible Region, 782
19.6 Grating Lobes, 784
19.7 Uniform Arrays, 786
19.8 Array Directivity, 790
19.9 Array Steering, 791
19.10 Array Beamwidth, 794
19.11 Problems, 796

20 Array Design Methods 799

20.1 Array Design Methods, 799
20.2 Schelkunoff's Zero Placement Method, 802
20.3 Fourier Series Method with Windowing, 804
20.4 Sector Beam Array Design, 805
20.5 Woodward-Lawson Frequency-Sampling Design, 809
20.6 Discretization of Continuous Line Sources, 814
20.7 Narrow-Beam Low-Sidelobe Designs, 818
20.8 Binomial Arrays, 822
20.9 Dolph-Chebyshev Arrays, 823
20.10 Taylor One-Parameter Source, 836
20.11 Prolate Array, 840
20.12 Taylor Line Source, 842
20.13 Villeneuve Arrays, 846
20.14 Multibeam Arrays, 847
20.15 Problems, 850

21 Currents on Linear Antennas 852

21.1 Hallén and Pocklington Integral Equations, 852
21.2 Delta-Gap, Frill Generator, and Plane-Wave Sources, 855
21.3 Solving Hallén's Equation, 856
21.4 Sinusoidal Current Approximation, 858
21.5 Reflecting and Center-Loaded Receiving Antennas, 859
21.6 King's Three-Term Approximation, 862
21.7 Evaluation of the Exact Kernel, 869
21.8 Method of Moments, 874
21.9 Delta-Function Basis, 877
21.10 Pulse Basis, 881
21.11 Triangular Basis, 886
21.12 NEC Sinusoidal Basis, 888
21.13 Hallén's Equation for Arbitrary Incident Field, 891
21.14 Solving Pocklington's Equation, 896
21.15 Problems, 900

22 Coupled Antennas 902

22.1 Near Fields of Linear Antennas, 902
22.2 Self and Mutual Impedance, 905
22.3 Coupled Two-Element Arrays, 911
22.4 Arrays of Parallel Dipoles, 914
22.5 Yagi-Uda Antennas, 923
22.6 Hallén Equations for Coupled Antennas, 929
22.7 Problems, 936

23 Appendices 938

A Physical Constants, 938
B Electromagnetic Frequency Bands, 939
C Vector Identities and Integral Theorems, 941
D Green's Functions, 944
E Coordinate Systems, 947
F Fresnel, Exponential, Sine, and Cosine Integrals, 949
G Gauss-Legendre Quadrature, 955
H Lorentz Transformations, 961
I MATLAB Functions, 969

References 974

Index 1021

Preface

This text provides a broad and applications-oriented introduction to electromagnetic waves and antennas. Current interest in these areas is driven by the growth in wireless and fiber-optic communications, information technology, and materials science.

Communications, antenna, radar, and microwave engineers must deal with the generation, transmission, and reception of electromagnetic waves. Device engineers working on ever-smaller integrated circuits and at ever higher frequencies must take into account wave propagation effects at the chip and circuit-board levels. Communication and computer network engineers routinely use waveguiding systems, such as transmission lines and optical fibers. Novel recent developments in materials, such as photonic bandgap structures, omnidirectional dielectric mirrors, birefringent multilayer films, surface plasmons, negative-index metamaterials, slow and fast light, promise a revolution in the control and manipulation of light and other applications. These are just some examples of topics discussed in this book. The text is organized around three main topic areas:

- The propagation, reflection, and transmission of plane waves, and the analysis and design of multilayer films.
- Waveguides, transmission lines, impedance matching, and S-parameters.
- Linear and aperture antennas, scalar and vector diffraction theory, antenna array design, numerical methods in antennas, and coupled antennas.

The text emphasizes connections to other subjects. For example, the mathematical techniques for analyzing wave propagation in multilayer structures and the design of multilayer optical filters are the same as those used in digital signal processing, such as the lattice structures of linear prediction, the analysis and synthesis of speech, and geophysical signal processing. Similarly, antenna array design is related to the problem of spectral analysis of sinusoids and to digital filter design, and Butler beams are equivalent to the FFT.

Use

The book is appropriate for first-year graduate or senior undergraduate students. There is enough material in the book for a two-semester course sequence. The book can also be used by practicing engineers and scientists who want a quick review that covers most of the basic concepts and includes many application examples.

The book is based on lecture notes for a first-year graduate course on “Electromagnetic Waves and Radiation” that I have been teaching at Rutgers for more than twenty years. The course draws students from a variety of fields, such as solid-state devices, wireless communications, fiber optics, biomedical engineering, and digital signal and array processing. Undergraduate seniors have also attended the graduate course successfully.

The book requires a prerequisite course on electromagnetics, typically offered at the junior year. Such introductory course is usually followed by a senior-level elective course on electromagnetic waves, which covers propagation, reflection, and transmission of waves, waveguides, transmission lines, and perhaps some antennas. This book may be used in such elective courses with the appropriate selection of chapters.

At the graduate level, there is usually an introductory course that covers waves, guides, lines, and antennas, and this is followed by more specialized courses on antenna design, microwave systems and devices, optical fibers, and numerical techniques in electromagnetics. No single book can possibly cover all of the advanced courses. This book may be used as a text in the initial course, and as a supplementary text in the specialized courses.

Contents and Highlights

The first eight chapters develop waves concepts and applications. The material progresses from Maxwell equations, to uniform plane waves in various media, such as lossless and lossy dielectrics and conductors, birefringent and chiral media, including negative-index media, to reflection and transmission problems at normal and oblique incidence, including reflection from moving boundaries and the Doppler effect, to multilayer structures.

Chapter three deals with pulse propagation in dispersive media, with discussions of group and front velocity and causality, group velocity dispersion, spreading and chirping, dispersion compensation, slow, fast, and negative group velocity, and an introduction to chirp radar and pulse compression.

Some of the oblique incidence applications include inhomogeneous waves, total internal reflection, surface plasmons, ray tracing and atmospheric refraction, and Snel’s law in negative-index media.

The material on multilayer structures includes the design of antireflection coatings, omnidirectional dielectric mirrors, broadband reflectionless multilayers, frustrated total internal reflection and surface plasmon resonance, perfect lenses in negative-index media, polarizing beam splitters, and birefringent multilayer structures.

Chapters 9–13 deal with waveguides and transmission lines. We cover only rectangular waveguides, resonant cavities, and simple dielectric waveguides. The transmission line material includes a discussion of microstrip and coaxial lines, terminated lines, standing wave ratio and the Smith chart, and examples of time-domain transient response of lines. We have included some material on coupled lines and crosstalk, as well as some on coupled mode theory and fiber Bragg gratings.

We devote one chapter to impedance matching methods, including multisection Chebyshev quarter-wavelength transformers, quarter-wavelength transformers with series or shunt stubs, single stub tuners, as well as *L*-section and *Π*-section reactive matching networks.

Chapter 13 presents an introduction to S-parameters with a discussion of input and output reflection coefficients, two-port stability conditions, transducer, operating, and available power gains, power waves, simultaneous conjugate matching, noise figure circles, illustrating the concepts with a number of low-noise high-gain microwave amplifier designs including the design of their input and output matching circuits.

Chapters 14–22 deal with radiation and antenna concepts. We begin by deriving expressions for the radiation fields from current sources, including magnetic currents, and then apply them to linear and aperture antennas. Chapter 15 covers general fundamental antenna concepts, such as radiation intensity, power density, directivity and gain, beamwidth, effective area, effective length, Friis formula, antenna noise temperature, power budgets in satellite links, and the radar equation.

We have included a number of linear antenna examples, such as Hertzian and half-wave dipoles, traveling, vee, and rhombic antennas, as well as loop antennas.

Two chapters are devoted to radiation from apertures. The first discusses Schelkunoff's field equivalence principle, magnetic currents and duality, radiation fields from apertures, vector diffraction theory, including the Kottler, Stratton-Chu, and Franz formulations, extinction theorem, Fresnel diffraction, Fresnel, zones, Sommerfeld's solution to the knife-edge diffraction problem, geometrical theory of diffraction, Rayleigh-Sommerfeld diffraction theory and its connection to the plane-wave spectrum representation with applications to Fourier optics.

The second presents a number of aperture antenna examples, such as open-ended waveguides, horn antennas, including optimum horn design, microstrip antennas, parabolic and dual reflectors, and lens antennas.

Two other chapters discuss antenna arrays. The first introduces basic concepts such as the multiplicative array pattern, visible region, grating lobes, directivity including its optimization, array steering, and beamwidth.

The other discusses several array design methods, such as by zero placement, Fourier series method with windowing, sector beam design, Woodward-Lawson method, and several narrow-beam low-sidelobe designs, such as binomial, Dolph-Chebyshev, Taylor's one-parameter, Taylor's $\bar{n}$ distribution, prolate, and Villeneuve array design. We have expanded on the analogies with time-domain DSP concepts and filter design methods. We finally give some examples of multibeam designs, such as Butler beams.

The last two chapters deal with numerical methods for linear antennas. Chapter 21 develops the Hallén and Pocklington integral equations for determining the current on a linear antenna, discusses King's three-term approximations, and then concentrates on numerical solutions for delta-gap input and arbitrary incident fields. We discuss the method of moments, implemented with the exact or the approximate thin-wire kernel and using various bases, such as pulse, triangular, and NEC bases. These methods require the accurate evaluation of the exact thin-wire kernel, which we approach using an elliptic function representation.

In Chapter 22 we discuss coupled antennas, in particular, parallel dipoles. Initially, we assume sinusoidal currents and reduce the problem to the calculation of the mutual impedance matrix. Then, we consider a more general formulation that requires the solution of a system of coupled Hallén equations. We present various examples, including the design of Yagi-Uda antennas.

Our MATLAB-based numerical solutions are not meant to replace sophisticated commercial field solvers. The inclusion of numerical methods in this book was motivated by the desire to provide the reader with some simple tools for self-study and experimentation. The study of numerical methods in electromagnetics is a subject in itself and our treatment does not do justice to it. However, we felt that it would be fun to be able to quickly compute fairly accurate radiation patterns in various antenna examples, such as Yagi-Uda and other coupled antennas, as well horns and reflector antennas.

The appendix includes summaries of physical constants, electromagnetic frequency bands, vector identities, integral theorems, Green's functions, coordinate systems, Fresnel integrals, sine and cosine integrals, the stationary phase approximation, Gauss-Legendre quadrature, Lorentz transformations, and a detailed list of the MATLAB functions.

Finally, there is a large (but inevitably incomplete) list of references, arranged by topic area, as well as several web links, that we hope could serve as a starting point for further study.

MATLAB Toolbox

The text makes extensive use of MATLAB. We have developed an "Electromagnetic Waves & Antennas" toolbox containing 170 MATLAB functions for carrying out all of the computations and simulation examples in the text. Code segments illustrating the usage of these functions are found throughout the book, and serve as a user manual. The functions may be grouped into the following categories:

1. Design and analysis of multilayer film structures, including antireflection coatings, polarizers, omnidirectional mirrors, narrow-band transmission filters, surface plasmon resonance, birefringent multilayer films and giant birefringent optics.
2. Design of quarter-wavelength impedance transformers and other impedance matching methods, such as Chebyshev transformers, dual-band transformers, stub matching and *L*-, *Π*- and *T*-section reactive matching networks.
3. Design and analysis of transmission lines and waveguides, such as microstrip lines and dielectric slab guides.
4. S-parameter functions for gain computations, Smith chart generation, stability, gain, and noise-figure circles, simultaneous conjugate matching, and microwave amplifier design.
5. Functions for the computation of directivities and gain patterns of linear antennas, such as dipole, vee, rhombic, and traveling-wave antennas, including functions for the input impedance of dipoles.
6. Aperture antenna functions for open-ended waveguides, horn antenna design, diffraction integrals, and knife-edge diffraction coefficients.
7. Antenna array design functions for uniform, binomial, Dolph-Chebyshev, Taylor one-parameter, Taylor $\bar{n}$ distribution, prolate, Villeneuve arrays, sector-beam,

multi-beam, Woodward-Lawson, and Butler beams. Functions for beamwidth and directivity calculations, and for steering and scanning arrays.

8. Numerical methods for solving the Hallén and Pocklington integral equations for single and coupled antennas, computing the exact thin-wire kernel, and computing self and mutual impedances.
9. Several functions for making azimuthal and polar plots of antenna and array gain patterns in decibels and absolute units.
10. There are also several MATLAB movies showing pulse propagation in dispersive media illustrating slow, fast, and negative group velocity; the propagation of step signals and pulses on terminated transmission lines; the propagation on cascaded lines; step signals getting reflected from reactive terminations; fault location by TDR; crosstalk signals propagating on coupled lines; and the time-evolution of the field lines radiated by a Hertzian dipole.

The MATLAB functions as well as other information about the book may be downloaded from the web page: `www.ece.rutgers.edu/~orfanidi/ewa`.

Acknowledgements

I would like to thank the many generations of my students who shaped the content of this book and the following people for their useful comments and suggestions for improvement: M. Abouowf, L. Alekseyev, S. Bang, R. Balder-Navarro, C. Christodoulou, C. Collister, A. Dane, G. Fano, H. Fluhler, K. Foster, S. Fuhrman, J. Hudson, F. Innes, W. G. Krische, M. Maybell, P. Matusov, K. T. McDonald, K. Michalski, J-S. Neron, V. Niziev, F. D. Nunes, H. Park, E. Perrin, D. Phillips, R. Rosensweig, M. Schuh, A. Siegman, P. Simon, K. Subramanian, V. Turkovic, P. Whiteneir, A. Young, and C. Zarowski. Any errors or shortcomings are, of course, entirely my own.

Sophocles J. Orfanidis
February 2008

Contents

Preface xiii

1 Maxwell's Equations 1

1.1 Maxwell's Equations, 1
1.2 Lorentz Force, 2
1.3 Constitutive Relations, 3
1.4 Boundary Conditions, 7
1.5 Currents, Fluxes, and Conservation Laws, 9
1.6 Charge Conservation, 10
1.7 Energy Flux and Energy Conservation, 11
1.8 Harmonic Time Dependence, 13
1.9 Simple Models of Dielectrics, Conductors, and Plasmas, 16
1.10 Kramers-Kronig Dispersion Relations, 26
1.11 Group Velocity, Energy Velocity, 29
1.12 Problems, 31

2 Uniform Plane Waves 36

2.1 Uniform Plane Waves in Lossless Media, 36
2.2 Monochromatic Waves, 42
2.3 Energy Density and Flux, 45
2.4 Wave Impedance, 46
2.5 Polarization, 46
2.6 Uniform Plane Waves in Lossy Media, 52
2.7 Propagation in Weakly Lossy Dielectrics, 58
2.8 Propagation in Good Conductors, 59
2.9 Propagation in Oblique Directions, 61
2.10 Complex or Inhomogeneous Waves, 63
2.11 Doppler Effect, 66
2.12 Propagation in Negative-Index Media, 70
2.13 Problems, 73

3 Pulse Propagation in Dispersive Media 80

3.1 Propagation Filter, 80
3.2 Front Velocity and Causality, 82
3.3 Exact Impulse Response Examples, 85
3.4 Transient and Steady-State Behavior, 88

3.5 Pulse Propagation and Group Velocity, 92
3.6 Group Velocity Dispersion and Pulse Spreading, 95
3.7 Propagation and Chirping, 100
3.8 Dispersion Compensation, 101
3.9 Slow, Fast, and Negative Group Velocities, 103
3.10 Chirp Radar and Pulse Compression, 110
3.11 Further Reading, 120
3.12 Problems, 120

4 Propagation in Birefringent Media 129

4.1 Linear and Circular Birefringence, 129
4.2 Uniaxial and Biaxial Media, 130
4.3 Chiral Media, 132
4.4 Gyrotropic Media, 135
4.5 Linear and Circular Dichroism, 136
4.6 Oblique Propagation in Birefringent Media, 137
4.7 Problems, 144

5 Reflection and Transmission 150

5.1 Propagation Matrices, 150
5.2 Matching Matrices, 154
5.3 Reflected and Transmitted Power, 157
5.4 Single Dielectric Slab, 160
5.5 Reflectionless Slab, 163
5.6 Time-Domain Reflection Response, 171
5.7 Two Dielectric Slabs, 173
5.8 Reflection by a Moving Boundary, 175
5.9 Problems, 178

6 Multilayer Structures 183

6.1 Multiple Dielectric Slabs, 183
6.2 Antireflection Coatings, 185
6.3 Dielectric Mirrors, 190
6.4 Propagation Bandgaps, 201
6.5 Narrow-Band Transmission Filters, 201
6.6 Equal Travel-Time Multilayer Structures, 206
6.7 Applications of Layered Structures, 220
6.8 Chebyshev Design of Reflectionless Multilayers, 223
6.9 Problems, 231

7 Oblique Incidence 238

7.1 Oblique Incidence and Snel's Laws, 238
7.2 Transverse Impedance, 240
7.3 Propagation and Matching of Transverse Fields, 243
7.4 Fresnel Reflection Coefficients, 245
7.5 Maximum Angle and Critical Angle, 247
7.6 Brewster Angle, 256

7.7 Complex Waves, 258
7.8 Total Internal Reflection, 261
7.9 Oblique Incidence on a Lossy Medium, 262
7.10 Zenneck Surface Wave, 267
7.11 Surface Plasmons, 269
7.12 Oblique Reflection from a Moving Boundary, 272
7.13 Geometrical Optics, 276
7.14 Fermat's Principle, 279
7.15 Ray Tracing, 281
7.16 Snel's Law in Negative-Index Media, 292
7.17 Problems, 295

8 Multilayer Film Applications 300

8.1 Multilayer Dielectric Structures at Oblique Incidence, 300
8.2 Lossy Multilayer Structures, 302
8.3 Single Dielectric Slab, 304
8.4 Frustrated Total Internal Reflection, 306
8.5 Surface Plasmon Resonance, 310
8.6 Perfect Lens in Negative-Index Media, 319
8.7 Antireflection Coatings at Oblique Incidence, 327
8.8 Omnidirectional Dielectric Mirrors, 330
8.9 Polarizing Beam Splitters, 341
8.10 Reflection and Refraction in Birefringent Media, 343
8.11 Brewster and Critical Angles in Birefringent Media, 347
8.12 Multilayer Birefringent Structures, 350
8.13 Giant Birefringent Optics, 352
8.14 Problems, 357

9 Waveguides 359

9.1 Longitudinal-Transverse Decompositions, 360
9.2 Power Transfer and Attenuation, 365
9.3 TEM, TE, and TM modes, 367
9.4 Rectangular Waveguides, 370
9.5 Higher TE and TM modes, 372
9.6 Operating Bandwidth, 374
9.7 Power Transfer, Energy Density, and Group Velocity, 375
9.8 Power Attenuation, 377
9.9 Reflection Model of Waveguide Propagation, 380
9.10 Resonant Cavities, 382
9.11 Dielectric Slab Waveguides, 384
9.12 Problems, 392

10 Transmission Lines 394

10.1 General Properties of TEM Transmission Lines, 394
10.2 Parallel Plate Lines, 400
10.3 Microstrip Lines, 401
10.4 Coaxial Lines, 405
10.5 Two-Wire Lines, 410

10.6 Distributed Circuit Model of a Transmission Line, 412
10.7 Wave Impedance and Reflection Response, 414
10.8 Two-Port Equivalent Circuit, 416
10.9 Terminated Transmission Lines, 417
10.10 Power Transfer from Generator to Load, 420
10.11 Open- and Short-Circuited Transmission Lines, 422
10.12 Standing Wave Ratio, 425
10.13 Determining an Unknown Load Impedance, 427
10.14 Smith Chart, 431
10.15 Time-Domain Response of Transmission Lines, 435
10.16 Problems, 442

11 Coupled Lines 453

11.1 Coupled Transmission Lines, 453
11.2 Crosstalk Between Lines, 459
11.3 Weakly Coupled Lines with Arbitrary Terminations, 462
11.4 Coupled-Mode Theory, 464
11.5 Fiber Bragg Gratings, 466
11.6 Diffuse Reflection and Transmission, 469
11.7 Problems, 471

12 Impedance Matching 473

12.1 Conjugate and Reflectionless Matching, 473
12.2 Multisection Transmission Lines, 475
12.3 Quarter-Wavelength Chebyshev Transformers, 476
12.4 Two-Section Dual-Band Chebyshev Transformers, 482
12.5 Quarter-Wavelength Transformer With Series Section, 488
12.6 Quarter-Wavelength Transformer With Shunt Stub, 491
12.7 Two-Section Series Impedance Transformer, 493
12.8 Single Stub Matching, 498
12.9 Balanced Stubs, 502
12.10 Double and Triple Stub Matching, 504
12.11 L-Section Lumped Reactive Matching Networks, 506
12.12 Pi-Section Lumped Reactive Matching Networks, 509
12.13 Reversed Matching Networks, 516
12.14 Problems, 518

13 S-Parameters 522

13.1 Scattering Parameters, 522
13.2 Power Flow, 526
13.3 Parameter Conversions, 527
13.4 Input and Output Reflection Coefficients, 528
13.5 Stability Circles, 530
13.6 Power Gains, 536
13.7 Generalized S-Parameters and Power Waves, 542
13.8 Simultaneous Conjugate Matching, 546
13.9 Power Gain Circles, 551
13.10 Unilateral Gain Circles, 552

13.11 Operating and Available Power Gain Circles, 554
13.12 Noise Figure Circles, 560
13.13 Problems, 565

14 Radiation Fields 568

14.1 Currents and Charges as Sources of Fields, 568
14.2 Retarded Potentials, 570
14.3 Harmonic Time Dependence, 573
14.4 Fields of a Linear Wire Antenna, 575
14.5 Fields of Electric and Magnetic Dipoles, 577
14.6 Ewald-Oseen Extinction Theorem, 582
14.7 Radiation Fields, 587
14.8 Radial Coordinates, 590
14.9 Radiation Field Approximation, 592
14.10 Computing the Radiation Fields, 593
14.11 Problems, 595

15 Transmitting and Receiving Antennas 598

15.1 Energy Flux and Radiation Intensity, 598
15.2 Directivity, Gain, and Beamwidth, 599
15.3 Effective Area, 604
15.4 Antenna Equivalent Circuits, 608
15.5 Effective Length, 610
15.6 Communicating Antennas, 612
15.7 Antenna Noise Temperature, 614
15.8 System Noise Temperature, 618
15.9 Data Rate Limits, 624
15.10 Satellite Links, 626
15.11 Radar Equation, 629
15.12 Problems, 631

16 Linear and Loop Antennas 634

16.1 Linear Antennas, 634
16.2 Hertzian Dipole, 636
16.3 Standing-Wave Antennas, 638
16.4 Half-Wave Dipole, 642
16.5 Monopole Antennas, 643
16.6 Traveling-Wave Antennas, 645
16.7 Vee and Rhombic Antennas, 647
16.8 Loop Antennas, 650
16.9 Circular Loops, 652
16.10 Square Loops, 654
16.11 Dipole and Quadrupole Radiation, 655
16.12 Problems, 657

17 Radiation from Apertures 658

17.1 Field Equivalence Principle, 658
17.2 Magnetic Currents and Duality, 660
17.3 Radiation Fields from Magnetic Currents, 662
17.4 Radiation Fields from Apertures, 663
17.5 Huygens Source, 666
17.6 Directivity and Effective Area of Apertures, 668
17.7 Uniform Apertures, 670
17.8 Rectangular Apertures, 670
17.9 Circular Apertures, 672
17.10 Vector Diffraction Theory, 675
17.11 Extinction Theorem, 679
17.12 Vector Diffraction for Apertures, 681
17.13 Fresnel Diffraction, 682
17.14 Knife-Edge Diffraction, 686
17.15 Geometrical Theory of Diffraction, 694
17.16 Rayleigh-Sommerfeld Diffraction Theory, 700
17.17 Plane-Wave Spectrum Representation, 703
17.18 Fresnel Diffraction and Fourier Optics, 708
17.19 Problems, 719

18 Aperture Antennas 723

18.1 Open-Ended Waveguides, 723
18.2 Horn Antennas, 727
18.3 Horn Radiation Fields, 729
18.4 Horn Directivity, 734
18.5 Horn Design, 737
18.6 Microstrip Antennas, 740
18.7 Parabolic Reflector Antennas, 746
18.8 Gain and Beamwidth of Reflector Antennas, 748
18.9 Aperture-Field and Current-Distribution Methods, 751
18.10 Radiation Patterns of Reflector Antennas, 754
18.11 Dual-Reflector Antennas, 763
18.12 Lens Antennas, 766

19 Antenna Arrays 768

19.1 Antenna Arrays, 768
19.2 Translational Phase Shift, 768
19.3 Array Pattern Multiplication, 770
19.4 One-Dimensional Arrays, 780
19.5 Visible Region, 782
19.6 Grating Lobes, 784
19.7 Uniform Arrays, 786
19.8 Array Directivity, 790
19.9 Array Steering, 791
19.10 Array Beamwidth, 794
19.11 Problems, 796

20 Array Design Methods 799

20.1 Array Design Methods, 799
20.2 Schelkunoff's Zero Placement Method, 802
20.3 Fourier Series Method with Windowing, 804
20.4 Sector Beam Array Design, 805
20.5 Woodward-Lawson Frequency-Sampling Design, 809
20.6 Discretization of Continuous Line Sources, 814
20.7 Narrow-Beam Low-Sidelobe Designs, 818
20.8 Binomial Arrays, 822
20.9 Dolph-Chebyshev Arrays, 823
20.10 Taylor One-Parameter Source, 836
20.11 Prolate Array, 840
20.12 Taylor Line Source, 842
20.13 Villeneuve Arrays, 846
20.14 Multibeam Arrays, 847
20.15 Problems, 850

21 Currents on Linear Antennas 852

21.1 Hallén and Pocklington Integral Equations, 852
21.2 Delta-Gap, Frill Generator, and Plane-Wave Sources, 855
21.3 Solving Hallén's Equation, 856
21.4 Sinusoidal Current Approximation, 858
21.5 Reflecting and Center-Loaded Receiving Antennas, 859
21.6 King's Three-Term Approximation, 862
21.7 Evaluation of the Exact Kernel, 869
21.8 Method of Moments, 874
21.9 Delta-Function Basis, 877
21.10 Pulse Basis, 881
21.11 Triangular Basis, 886
21.12 NEC Sinusoidal Basis, 888
21.13 Hallén's Equation for Arbitrary Incident Field, 891
21.14 Solving Pocklington's Equation, 896
21.15 Problems, 900

22 Coupled Antennas 902

22.1 Near Fields of Linear Antennas, 902
22.2 Self and Mutual Impedance, 905
22.3 Coupled Two-Element Arrays, 911
22.4 Arrays of Parallel Dipoles, 914
22.5 Yagi-Uda Antennas, 923
22.6 Hallén Equations for Coupled Antennas, 929
22.7 Problems, 936

23 Appendices 938

A Physical Constants, 938
B Electromagnetic Frequency Bands, 939
C Vector Identities and Integral Theorems, 941
D Green's Functions, 944
E Coordinate Systems, 947
F Fresnel, Exponential, Sine, and Cosine Integrals, 949
G Gauss-Legendre Quadrature, 955
H Lorentz Transformations, 961
I MATLAB Functions, 969

References 974

Index 1021

1

Maxwell's Equations

1.1 Maxwell's Equations

Maxwell's equations describe all (classical) electromagnetic phenomena:

$$\boxed{\begin{aligned} \nabla\times\boldsymbol{E} &= -\frac{\partial\boldsymbol{B}}{\partial t} \\ \nabla\times\boldsymbol{H} &= \boldsymbol{J}+\frac{\partial\boldsymbol{D}}{\partial t} \\ \nabla\cdot\boldsymbol{D} &= \rho \\ \nabla\cdot\boldsymbol{B} &= 0 \end{aligned}} \qquad \text{(Maxwell's equations)} \tag{1.1.1}$$

The first is *Faraday's law of induction*, the second is *Ampère's law* as amended by Maxwell to include the displacement current $\partial\boldsymbol{D}/\partial t$, the third and fourth are *Gauss' laws* for the electric and magnetic fields.

The displacement current term $\partial\boldsymbol{D}/\partial t$ in Ampère's law is essential in predicting the existence of propagating electromagnetic waves. Its role in establishing charge conservation is discussed in Sec. 1.6.

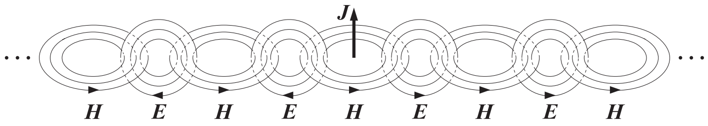
Eqs. (1.1.1) are in SI units. The quantities $\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and magnetic *field intensities* and are measured in units of [volt/m] and [ampere/m], respectively. The quantities $\boldsymbol{D}$ and $\boldsymbol{B}$ are the electric and magnetic *flux densities* and are in units of [coulomb/m^2] and [weber/m^2], or [tesla]. $\boldsymbol{B}$ is also called the *magnetic induction*.

The quantities ρ and $\boldsymbol{J}$ are the *volume charge density* and *electric current density* (charge flux) of any *external* charges (that is, not including any induced polarization charges and currents.) They are measured in units of [coulomb/m^3] and [ampere/m^2]. The right-hand side of the fourth equation is zero because there are no magnetic monopole charges. Eqs. (1.3.17)–(1.3.19) display the induced polarization terms explicitly.

The charge and current densities $\rho, \boldsymbol{J}$ may be thought of as the *sources* of the electromagnetic fields. For wave propagation problems, these densities are localized in space; for example, they are restricted to flow on an antenna. The generated electric and magnetic fields are *radiated* away from these sources and can propagate to large distances to the receiving antennas. Away from the sources, that is, in source-free regions of space, Maxwell's equations take the simpler form:

$$\boxed{\begin{aligned} \nabla\times\boldsymbol{E} &= -\frac{\partial\boldsymbol{B}}{\partial t}\\ \nabla\times\boldsymbol{H} &= \frac{\partial\boldsymbol{D}}{\partial t}\\ \nabla\cdot\boldsymbol{D} &= 0\\ \nabla\cdot\boldsymbol{B} &= 0 \end{aligned}} \qquad \text{(source-free Maxwell's equations)} \tag{1.1.2}$$

The qualitative mechanism by which Maxwell's equations give rise to propagating electromagnetic fields is shown in the figure below.



For example, a time-varying current $\boldsymbol{J}$ on a linear antenna generates a circulating and time-varying magnetic field $\boldsymbol{H}$, which through Faraday's law generates a circulating electric field $\boldsymbol{E}$, which through Ampère's law generates a magnetic field, and so on. The cross-linked electric and magnetic fields propagate away from the current source. A more precise discussion of the fields radiated by a localized current distribution is given in Chap. 14.

1.2 Lorentz Force

The force on a charge q moving with velocity $\boldsymbol{v}$ in the presence of an electric and magnetic field $\boldsymbol{E},\boldsymbol{B}$ is called the Lorentz force and is given by:

$$\boldsymbol{F} = q\,(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}) \qquad \text{(Lorentz force)} \tag{1.2.1}$$

Newton's equation of motion is (for non-relativistic speeds):

$$m\,\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F} = q\,(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}) \tag{1.2.2}$$

where m is the mass of the charge. The force $\boldsymbol{F}$ will increase the kinetic energy of the charge at a rate that is equal to the rate of work done by the Lorentz force on the charge, that is, $\boldsymbol{v}\cdot\boldsymbol{F}$. Indeed, the time-derivative of the kinetic energy is:

$$W_{\text{kin}} = \frac{1}{2}m\,\boldsymbol{v}\cdot\boldsymbol{v} \quad\Rightarrow\quad \frac{dW_{\text{kin}}}{dt} = m\,\boldsymbol{v}\cdot\frac{d\boldsymbol{v}}{dt} = \boldsymbol{v}\cdot\boldsymbol{F} = q\,\boldsymbol{v}\cdot\boldsymbol{E} \tag{1.2.3}$$

We note that only the electric force contributes to the increase of the kinetic energy—the magnetic force remains perpendicular to $\boldsymbol{v}$, that is, $\boldsymbol{v}\cdot(\boldsymbol{v}\times\boldsymbol{B}) = 0$.

Volume charge and current distributions $\rho,\boldsymbol{J}$ are also subjected to forces in the presence of fields. The Lorentz force *per unit volume* acting on $\rho,\boldsymbol{J}$ is given by:

$$\boldsymbol{f} = \rho\boldsymbol{E}+\boldsymbol{J}\times\boldsymbol{B} \qquad \text{(Lorentz force per unit volume)} \tag{1.2.4}$$

where $\boldsymbol{f}$ is measured in units of [newton/m^3]. If $\boldsymbol{J}$ arises from the motion of charges within the distribution ρ, then $\boldsymbol{J} = \rho\boldsymbol{v}$ (as explained in Sec. 1.5.) In this case,

$$\boldsymbol{f} = \rho\,(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}) \tag{1.2.5}$$

By analogy with Eq. (1.2.3), the quantity $\boldsymbol{v}\cdot\boldsymbol{f} = \rho\,\boldsymbol{v}\cdot\boldsymbol{E} = \boldsymbol{J}\cdot\boldsymbol{E}$ represents the *power per unit volume* of the forces acting on the moving charges, that is, the power expended by (or lost from) the fields and converted into kinetic energy of the charges, or heat. It has units of [watts/m^3]. We will denote it by:

$$\boxed{\frac{dP_{\text{loss}}}{dV} = \boldsymbol{J}\cdot\boldsymbol{E}} \qquad \text{(ohmic power losses per unit volume)} \tag{1.2.6}$$

In Sec. 1.7, we discuss its role in the conservation of energy. We will find that electromagnetic energy flowing into a region will partially increase the stored energy in that region and partially dissipate into heat according to Eq. (1.2.6).

1.3 Constitutive Relations

The electric and magnetic flux densities $\boldsymbol{D},\boldsymbol{B}$ are related to the field intensities $\boldsymbol{E},\boldsymbol{H}$ via the so-called *constitutive relations*, whose precise form depends on the material in which the fields exist. In *vacuum*, they take their simplest form:

$$\boxed{\begin{aligned} \boldsymbol{D} &= \epsilon_0\boldsymbol{E}\\ \boldsymbol{B} &= \mu_0\boldsymbol{H} \end{aligned}} \tag{1.3.1}$$

where ϵ_0,μ_0 are the *permittivity* and *permeability* of vacuum, with numerical values:

$$\boxed{\begin{aligned} \epsilon_0 &= 8.854\times10^{-12}\ \text{farad/m}\\ \mu_0 &= 4\pi\times10^{-7}\ \text{henry/m} \end{aligned}} \tag{1.3.2}$$

The units for ϵ_0 and μ_0 are the units of the ratios D/E and B/H, that is,

$$\frac{\text{coulomb/m}^2}{\text{volt/m}} = \frac{\text{coulomb}}{\text{volt}\cdot\text{m}} = \frac{\text{farad}}{\text{m}}\,, \qquad \frac{\text{weber/m}^2}{\text{ampere/m}} = \frac{\text{weber}}{\text{ampere}\cdot\text{m}} = \frac{\text{henry}}{\text{m}}$$

From the two quantities ϵ_0,μ_0, we can define two other physical constants, namely, the *speed of light* and *characteristic impedance* of vacuum:

$$\boxed{c_0 = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 3\times10^8\ \text{m/sec}\,, \qquad \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377\ \text{ohm}} \tag{1.3.3}$$

The next simplest form of the constitutive relations is for simple homogeneous isotropic dielectric and for magnetic materials:

$$\boxed{\begin{aligned} \boldsymbol{D} &= \epsilon \boldsymbol{E} \\ \boldsymbol{B} &= \mu \boldsymbol{H} \end{aligned}} \tag{1.3.4}$$

These are typically valid at low frequencies. The permittivity ϵ and permeability μ are related to the *electric and magnetic susceptibilities* of the material as follows:

$$\boxed{\begin{aligned} \epsilon &= \epsilon_0(1+\chi) \\ \mu &= \mu_0(1+\chi_m) \end{aligned}} \tag{1.3.5}$$

The susceptibilities χ, χ_m are measures of the electric and magnetic polarization properties of the material. For example, we have for the electric flux density:

$$\boldsymbol{D} = \epsilon \boldsymbol{E} = \epsilon_0(1+\chi)\boldsymbol{E} = \epsilon_0 \boldsymbol{E} + \epsilon_0 \chi \boldsymbol{E} = \epsilon_0 \boldsymbol{E} + \boldsymbol{P} \tag{1.3.6}$$

where the quantity $\boldsymbol{P} = \epsilon_0 \chi \boldsymbol{E}$ represents the dielectric polarization of the material, that is, the average electric dipole moment per unit volume. In a magnetic material, we have

$$\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M}) = \mu_0(\boldsymbol{H} + \chi_m \boldsymbol{H}) = \mu_0(1+\chi_m)\boldsymbol{H} = \mu \boldsymbol{H} \tag{1.3.7}$$

where $\boldsymbol{M} = \chi_m \boldsymbol{H}$ is the *magnetization*, that is, the average magnetic moment per unit volume. The speed of light in the material and the characteristic impedance are:

$$c = \frac{1}{\sqrt{\mu\epsilon}}\,, \qquad \eta = \sqrt{\frac{\mu}{\epsilon}} \tag{1.3.8}$$

The *relative permittivity, permeability* and *refractive index* of a material are defined by:

$$\epsilon_{\text{rel}} = \frac{\epsilon}{\epsilon_0} = 1 + \chi\,, \qquad \mu_{\text{rel}} = \frac{\mu}{\mu_0} = 1 + \chi_m\,, \qquad n = \sqrt{\epsilon_{\text{rel}}\mu_{\text{rel}}} \tag{1.3.9}$$

so that $n^2 = \epsilon_{\text{rel}}\mu_{\text{rel}}$. Using the definition of Eq. (1.3.8), we may relate the speed of light and impedance of the material to the corresponding vacuum values:

$$\begin{aligned} c &= \frac{1}{\sqrt{\mu\epsilon}} = \frac{1}{\sqrt{\mu_0\epsilon_0\epsilon_{\text{rel}}\mu_{\text{rel}}}} = \frac{c_0}{\sqrt{\epsilon_{\text{rel}}\mu_{\text{rel}}}} = \frac{c_0}{n} \\ \eta &= \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_0}{\epsilon_0}}\sqrt{\frac{\mu_{\text{rel}}}{\epsilon_{\text{rel}}}} = \eta_0\sqrt{\frac{\mu_{\text{rel}}}{\epsilon_{\text{rel}}}} = \eta_0\,\frac{\mu_{\text{rel}}}{n} = \eta_0\,\frac{n}{\epsilon_{\text{rel}}} \end{aligned} \tag{1.3.10}$$

For a non-magnetic material, we have $\mu = \mu_0$, or, $\mu_{\text{rel}} = 1$, and the impedance becomes simply $\eta = \eta_0/n$, a relationship that we will use extensively in this book.

More generally, constitutive relations may be inhomogeneous, anisotropic, nonlinear, frequency dependent (dispersive), or all of the above. In *inhomogeneous materials*, the permittivity ϵ depends on the location within the material:

$$\boldsymbol{D}(\mathbf{r}, t) = \epsilon(\mathbf{r})\boldsymbol{E}(\mathbf{r}, t)$$

In *anisotropic materials*, ϵ depends on the x, y, z direction and the constitutive relations may be written component-wise in matrix (or tensor) form:

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{1.3.11}$$

Anisotropy is an inherent property of the atomic/molecular structure of the dielectric. It may also be caused by the application of external fields. For example, conductors and plasmas in the presence of a constant magnetic field—such as the ionosphere in the presence of the Earth's magnetic field—become anisotropic (see for example, Problem 1.10 on the Hall effect.)

In *nonlinear materials*, ϵ may depend on the magnitude E of the applied electric field in the form:

$$\boldsymbol{D} = \epsilon(E)\boldsymbol{E}\,, \qquad \text{where} \qquad \epsilon(E) = \epsilon + \epsilon_2 E + \epsilon_3 E^2 + \cdots \tag{1.3.12}$$

Nonlinear effects are desirable in some applications, such as various types of electro-optic effects used in light phase modulators and phase retarders for altering polarization. In other applications, however, they are undesirable. For example, in optical fibers nonlinear effects become important if the transmitted power is increased beyond a few milliwatts. A typical consequence of nonlinearity is to cause the generation of higher harmonics, for example, if $E = E_0 e^{j\omega t}$, then Eq. (1.3.12) gives:

$$D = \epsilon(E)E = \epsilon E + \epsilon_2 E^2 + \epsilon_3 E^3 + \cdots = \epsilon E_0 e^{j\omega t} + \epsilon_2 E_0^2 e^{2j\omega t} + \epsilon_3 E_0^3 e^{3j\omega t} + \cdots$$

Thus the input frequency ω is replaced by $\omega, 2\omega, 3\omega$, and so on. In a multi-wavelength transmission system, such as a wavelength division multiplexed (WDM) optical fiber system carrying signals at closely-spaced carrier frequencies, such nonlinearities will cause the appearance of new frequencies which may be viewed as crosstalk among the original channels. For example, if the system carries frequencies ω_i, $i = 1, 2, \dots$, then the presence of a cubic nonlinearity E^3 will cause the appearance of the frequencies $\omega_i \pm \omega_j \pm \omega_k$. In particular, the frequencies $\omega_i + \omega_j - \omega_k$ are most likely to be confused as crosstalk because of the close spacing of the carrier frequencies.

Materials with a *frequency-dependent* dielectric constant $\epsilon(\omega)$ are referred to as *dispersive*. The frequency dependence comes about because when a time-varying electric field is applied, the polarization response of the material cannot be instantaneous. Such *dynamic* response can be described by the convolutional (and causal) constitutive relationship:

$$\boldsymbol{D}(\mathbf{r}, t) = \int_{-\infty}^{t} \epsilon(t - t')\boldsymbol{E}(\mathbf{r}, t')\,dt' \tag{1.3.13}$$

which becomes multiplicative in the frequency domain:

$$\boxed{\boldsymbol{D}(\mathbf{r}, \omega) = \epsilon(\omega)\boldsymbol{E}(\mathbf{r}, \omega)} \tag{1.3.14}$$

All materials are, in fact, dispersive. However, $\epsilon(\omega)$ typically exhibits strong dependence on ω only for certain frequencies. For example, water at optical frequencies has refractive index $n = \sqrt{\epsilon_{\text{rel}}} = 1.33$, but at RF down to dc, it has $n = 9$.

In Sec. 1.9, we discuss simple models of $\epsilon(\omega)$ for dielectrics, conductors, and plasmas, and clarify the nature of Ohm's law:

$$\boxed{\boldsymbol{J} = \sigma \boldsymbol{E}} \qquad \text{(Ohm's law)} \tag{1.3.15}$$

In Sec. 1.10, we discuss the Kramers-Kronig dispersion relations, which are a direct consequence of the causality of the time-domain dielectric response function $\epsilon(t)$.

One major consequence of material dispersion is *pulse spreading*, that is, the progressive widening of a pulse as it propagates through such a material. This effect limits the data rate at which pulses can be transmitted. There are other types of dispersion, such as *intermodal dispersion* in which several modes may propagate simultaneously, or *waveguide dispersion* introduced by the confining walls of a waveguide.

There exist materials that are both nonlinear and dispersive that support certain types of non-linear waves called *solitons*, in which the spreading effect of dispersion is exactly canceled by the nonlinearity. Therefore, soliton pulses maintain their shape as they propagate in such media [1151,850,851].

More complicated forms of constitutive relationships arise in chiral and gyrotropic media and are discussed in Chap. 4. The more general bi-isotropic and bi-anisotropic media are discussed in [30,73].

In Eqs. (1.1.1), the densities $\rho, \boldsymbol{J}$ represent the *external* or *free* charges and currents in a material medium. The induced polarization $\boldsymbol{P}$ and magnetization $\boldsymbol{M}$ may be made explicit in Maxwell's equations by using constitutive relations:

$$\boldsymbol{D} = \epsilon_0 \boldsymbol{E} + \boldsymbol{P}, \quad \boldsymbol{B} = \mu_0 (\boldsymbol{H} + \boldsymbol{M}) \tag{1.3.16}$$

Inserting these in Eq. (1.1.1), for example, by writing $\nabla\times\boldsymbol{B} = \mu_0 \nabla\times(\boldsymbol{H}+\boldsymbol{M}) = \mu_0(\boldsymbol{J} + \dot{\boldsymbol{D}} + \nabla\times\boldsymbol{M}) = \mu_0(\epsilon_0\dot{\boldsymbol{E}} + \boldsymbol{J} + \dot{\boldsymbol{P}} + \nabla\times\boldsymbol{M})$, we may express Maxwell's equations in terms of the fields $\boldsymbol{E}$ and $\boldsymbol{B}$:

$$\begin{aligned}
\nabla\times\boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t} \\
\nabla\times\boldsymbol{B} &= \mu_0\epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \mu_0\Big[\boldsymbol{J} + \frac{\partial \boldsymbol{P}}{\partial t} + \nabla\times\boldsymbol{M}\Big] \\
\nabla\cdot\boldsymbol{E} &= \frac{1}{\epsilon_0}(\rho - \nabla\cdot\boldsymbol{P}) \\
\nabla\cdot\boldsymbol{B} &= 0
\end{aligned} \tag{1.3.17}$$

We identify the current and charge densities due to the polarization of the material as:

$$\boxed{\boldsymbol{J}_{\text{pol}} = \frac{\partial \boldsymbol{P}}{\partial t}, \quad \rho_{\text{pol}} = -\nabla\cdot\boldsymbol{P}} \qquad \text{(polarization densities)} \tag{1.3.18}$$

Similarly, the quantity $\boldsymbol{J}_{\text{mag}} = \nabla\times\boldsymbol{M}$ may be identified as the magnetization current density (note that $\rho_{\text{mag}} = 0$.) The total current and charge densities are:

$$\begin{aligned}
\boldsymbol{J}_{\text{tot}} &= \boldsymbol{J} + \boldsymbol{J}_{\text{pol}} + \boldsymbol{J}_{\text{mag}} = \boldsymbol{J} + \frac{\partial \boldsymbol{P}}{\partial t} + \nabla\times\boldsymbol{M} \\
\rho_{\text{tot}} &= \rho + \rho_{\text{pol}} = \rho - \nabla\cdot\boldsymbol{P}
\end{aligned} \tag{1.3.19}$$

and may be thought of as the *sources* of the fields in Eq. (1.3.17). In Sec. 14.6, we examine this interpretation further and show how it leads to the Ewald-Oseen extinction theorem and to a microscopic explanation of the origin of the refractive index.

Negative Index Media

Maxwell's equations do not preclude the possibility that one or both of the quantities ϵ, μ be negative. For example, plasmas below their plasma frequency, and metals up to optical frequencies, have $\epsilon < 0$ and $\mu > 0$, with interesting applications such as surface plasmons (see Sec. 8.5).

Isotropic media with $\mu < 0$ and $\epsilon > 0$ are more difficult to come by [130], although examples of such media have been fabricated [358].

Negative-index media, also known as left-handed media, have ϵ, μ that are simultaneously negative, $\epsilon < 0$ and $\mu < 0$. Veselago [353] was the first to study their unusual electromagnetic properties, such as having a negative index of refraction and the reversal of Snel's law.

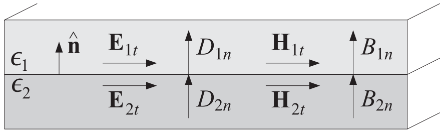
The novel properties of such media and their potential applications have generated a lot of research interest [353–434]. Examples of such media, termed "metamaterials", have been constructed using periodic arrays of wires and split-ring resonators, [359] and by transmission line elements [392–394,414,427], and have been shown to exhibit the properties predicted by Veselago.

When $\epsilon_{\text{rel}} < 0$ and $\mu_{\text{rel}} < 0$, the refractive index, $n^2 = \epsilon_{\text{rel}}\mu_{\text{rel}}$, must be defined by the negative square root $n = -\sqrt{\epsilon_{\text{rel}}\mu_{\text{rel}}}$. Because then $n < 0$ and $\mu_{\text{rel}} < 0$ will imply that the characteristic impedance of the medium $\eta = \eta_0\mu_{\text{rel}}/n$ will be positive, which as we will see later implies that the energy flux of a wave is in the same direction as the direction of propagation. We discuss such media in Sections 2.12, 7.16, and 8.6.

1.4 Boundary Conditions

The boundary conditions for the electromagnetic fields across material boundaries are given below:

$$\boxed{\begin{aligned}
\boldsymbol{E}_{1t} - \boldsymbol{E}_{2t} &= 0 \\
\boldsymbol{H}_{1t} - \boldsymbol{H}_{2t} &= \mathbf{J}_s \times \hat{\mathbf{n}} \\
D_{1n} - D_{2n} &= \rho_s \\
B_{1n} - B_{2n} &= 0
\end{aligned}} \tag{1.4.1}$$



where $\hat{\mathbf{n}}$ is a unit vector normal to the boundary pointing from medium-2 into medium-1. The quantities $\rho_s, \mathbf{J}_s$ are any external *surface charge* and *surface current* densities on the boundary surface and are measured in units of [coulomb/m^2] and [ampere/m].

In words, the *tangential* components of the $\boldsymbol{E}$-field are continuous across the interface; the difference of the *tangential* components of the $\boldsymbol{H}$-field are equal to the surface current density; the difference of the *normal* components of the flux density $\boldsymbol{D}$ are equal to the surface charge density; and the *normal* components of the magnetic flux density $\boldsymbol{B}$ are continuous.

The D_n boundary condition may also be written a form that brings out the dependence on the polarization surface charges:

$$(\epsilon_0 E_{1n}+P_{1n})-(\epsilon_0 E_{2n}+P_{2n})=\rho_s \quad\Rightarrow\quad \epsilon_0(E_{1n}-E_{2n})=\rho_s-P_{1n}+P_{2n}=\rho_{s,\text{tot}}$$

The total surface charge density will be $\rho_{s,\text{tot}}=\rho_s+\rho_{1s,\text{pol}}+\rho_{2s,\text{pol}}$, where the surface charge density of polarization charges accumulating at the surface of a dielectric is seen to be ($\hat{\mathbf{n}}$ is the outward normal from the dielectric):

$$\boxed{\rho_{s,\text{pol}}=P_n=\hat{\mathbf{n}}\cdot\boldsymbol{P}} \tag{1.4.2}$$

The relative directions of the field vectors are shown in Fig. 1.4.1. Each vector may be decomposed as the sum of a part tangential to the surface and a part perpendicular to it, that is, $\boldsymbol{E}=\boldsymbol{E}_t+\boldsymbol{E}_n$. Using the vector identity,

$$\boldsymbol{E}=\hat{\mathbf{n}}\times(\boldsymbol{E}\times\hat{\mathbf{n}})+\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\boldsymbol{E})=\boldsymbol{E}_t+\boldsymbol{E}_n \tag{1.4.3}$$

we identify these two parts as:

$$\boldsymbol{E}_t=\hat{\mathbf{n}}\times(\boldsymbol{E}\times\hat{\mathbf{n}})\,,\qquad \boldsymbol{E}_n=\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\boldsymbol{E})=\hat{\mathbf{n}}E_n$$

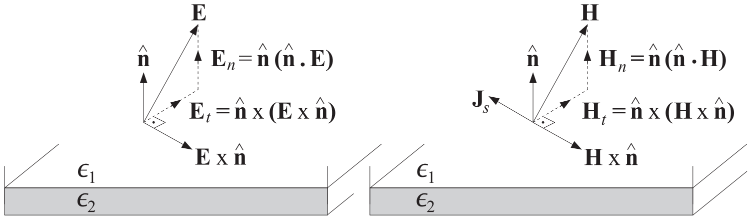


Fig. 1.4.1 Field directions at boundary.

Using these results, we can write the first two boundary conditions in the following vectorial forms, where the second form is obtained by taking the cross product of the first with $\hat{\mathbf{n}}$ and noting that $\boldsymbol{J}_s$ is purely tangential:

$$\boxed{\begin{aligned}\hat{\mathbf{n}}\times(\boldsymbol{E}_1\times\hat{\mathbf{n}})-\hat{\mathbf{n}}\times(\boldsymbol{E}_2\times\hat{\mathbf{n}})&=0\\ \hat{\mathbf{n}}\times(\boldsymbol{H}_1\times\hat{\mathbf{n}})-\hat{\mathbf{n}}\times(\boldsymbol{H}_2\times\hat{\mathbf{n}})&=\boldsymbol{J}_s\times\hat{\mathbf{n}}\end{aligned}}\quad\text{or,}\quad\boxed{\begin{aligned}\hat{\mathbf{n}}\times(\boldsymbol{E}_1-\boldsymbol{E}_2)&=0\\ \hat{\mathbf{n}}\times(\boldsymbol{H}_1-\boldsymbol{H}_2)&=\boldsymbol{J}_s\end{aligned}} \tag{1.4.4}$$

The boundary conditions (1.4.1) can be derived from the integrated form of Maxwell's equations if we make some additional regularity assumptions about the fields at the interfaces.

In many interface problems, there are no externally applied surface charges or currents on the boundary. In such cases, the boundary conditions may be stated as:

$$\boxed{\begin{aligned}\boldsymbol{E}_{1t}&=\boldsymbol{E}_{2t}\\ \boldsymbol{H}_{1t}&=\boldsymbol{H}_{2t}\\ D_{1n}&=D_{2n}\\ B_{1n}&=B_{2n}\end{aligned}}\qquad\text{(source-free boundary conditions)} \tag{1.4.5}$$

1.5 Currents, Fluxes, and Conservation Laws

The electric current density $\boldsymbol{J}$ is an example of a *flux vector* representing the flow of the electric charge. The concept of flux is more general and applies to any quantity that flows.[†] It could, for example, apply to energy flux, momentum flux (which translates into pressure force), mass flux, and so on.

In general, the flux of a quantity Q is defined as the amount of the quantity that flows (perpendicularly) through a unit surface in unit time. Thus, if the amount ΔQ flows through the surface ΔS in time Δt, then:

$$J=\frac{\Delta Q}{\Delta S\Delta t}\qquad\text{(definition of flux)} \tag{1.5.1}$$

When the flowing quantity Q is the electric charge, the amount of current through the surface ΔS will be $\Delta I=\Delta Q/\Delta t$, and therefore, we can write $J=\Delta I/\Delta S$, with units of [ampere/m^2].

The flux is a vectorial quantity whose direction points in the direction of flow. There is a fundamental relationship that relates the flux vector $\boldsymbol{J}$ to the transport velocity $\boldsymbol{v}$ and the volume density ρ of the flowing quantity:

$$\boxed{\boldsymbol{J}=\rho\boldsymbol{v}} \tag{1.5.2}$$

This can be derived with the help of Fig. 1.5.1. Consider a surface ΔS oriented perpendicularly to the flow velocity. In time Δt, the entire amount of the quantity contained in the cylindrical volume of height $v\Delta t$ will manage to flow through ΔS. This amount is equal to the density of the material times the cylindrical volume $\Delta V=\Delta S(v\Delta t)$, that is, $\Delta Q=\rho\Delta V=\rho\,\Delta S\,v\Delta t$. Thus, by definition:

$$J=\frac{\Delta Q}{\Delta S\Delta t}=\frac{\rho\,\Delta S\,v\Delta t}{\Delta S\Delta t}=\rho v$$

When $\boldsymbol{J}$ represents electric current density, we will see in Sec. 1.9 that Eq. (1.5.2) implies Ohm's law $\boldsymbol{J}=\sigma\boldsymbol{E}$. When the vector $\boldsymbol{J}$ represents the energy flux of a propagating electromagnetic wave and ρ the corresponding energy per unit volume, then because the speed of propagation is the velocity of light, we expect that Eq. (1.5.2) will take the form:

$$J_{\text{en}}=c\rho_{\text{en}} \tag{1.5.3}$$

[†] In this sense, the terms electric and magnetic "flux densities" for the quantities $\boldsymbol{D},\boldsymbol{B}$ are somewhat of a misnomer because they do not represent anything that flows.

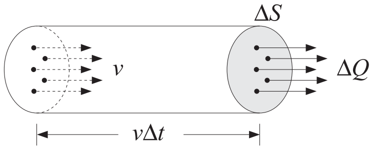


Fig. 1.5.1 Flux of a quantity.

Similarly, when J represents momentum flux, we expect to have $J_{\rm mom} = c\rho_{\rm mom}$. Momentum flux is defined as $J_{\rm mom} = \Delta p/(\Delta S\Delta t) = \Delta F/\Delta S$, where p denotes momentum and $\Delta F = \Delta p/\Delta t$ is the rate of change of momentum, or the force, exerted on the surface ΔS. Thus, $J_{\rm mom}$ represents force per unit area, or pressure.

Electromagnetic waves incident on material surfaces exert pressure (known as radiation pressure), which can be calculated from the momentum flux vector. It can be shown that the momentum flux is numerically equal to the energy density of a wave, that is, $J_{\rm mom} = \rho_{\rm en}$, which implies that $\rho_{\rm en} = \rho_{\rm mom}c$. This is consistent with the theory of relativity, which states that the energy-momentum relationship for a photon is $E = pc$.

1.6 Charge Conservation

Maxwell added the displacement current term to Ampère's law in order to guarantee charge conservation. Indeed, taking the divergence of both sides of Ampère's law and using Gauss's law $\boldsymbol{\nabla}\cdot\boldsymbol{D} = \rho$, we get:

$$\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\boldsymbol{H} = \boldsymbol{\nabla}\cdot\boldsymbol{J} + \boldsymbol{\nabla}\cdot\frac{\partial\boldsymbol{D}}{\partial t} = \boldsymbol{\nabla}\cdot\boldsymbol{J} + \frac{\partial}{\partial t}\boldsymbol{\nabla}\cdot\boldsymbol{D} = \boldsymbol{\nabla}\cdot\boldsymbol{J} + \frac{\partial\rho}{\partial t}$$

Using the vector identity $\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}\times\boldsymbol{H} = 0$, we obtain the differential form of the charge conservation law:

$$\boxed{\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{J} = 0} \qquad \text{(charge conservation)} \tag{1.6.1}$$

Integrating both sides over a closed volume V surrounded by the surface S, as shown in Fig. 1.6.1, and using the divergence theorem, we obtain the integrated form of Eq. (1.6.1):

$$\oint_S \boldsymbol{J}\cdot d\boldsymbol{S} = -\frac{d}{dt}\int_V \rho\, dV \tag{1.6.2}$$

The left-hand side represents the total amount of charge flowing *outwards* through the surface S per unit time. The right-hand side represents the amount by which the charge is *decreasing* inside the volume V per unit time. In other words, charge does not disappear into (or get created out of) nothingness—it decreases in a region of space only because it flows into other regions.

Another consequence of Eq. (1.6.1) is that in good conductors, there cannot be any accumulated *volume* charge. Any such charge will quickly move to the conductor's surface and distribute itself such that to make the surface into an equipotential surface.

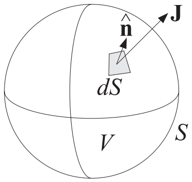


Fig. 1.6.1 Flux outwards through surface.

Assuming that inside the conductor we have $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{J} = \sigma\boldsymbol{E}$, we obtain

$$\boldsymbol{\nabla}\cdot\boldsymbol{J} = \sigma\boldsymbol{\nabla}\cdot\boldsymbol{E} = \frac{\sigma}{\epsilon}\boldsymbol{\nabla}\cdot\boldsymbol{D} = \frac{\sigma}{\epsilon}\rho$$

$$\frac{\partial\rho}{\partial t} + \frac{\sigma}{\epsilon}\rho = 0 \tag{1.6.3}$$

with solution:

$$\rho(\mathbf{r},t) = \rho_0(\mathbf{r})\,e^{-\sigma t/\epsilon}$$

where $\rho_0(\mathbf{r})$ is the initial volume charge distribution. The solution shows that the volume charge disappears from inside and therefore it must accumulate on the surface of the conductor. The "relaxation" time constant $\tau_{\rm rel} = \epsilon/\sigma$ is extremely short for good conductors. For example, in copper,

$$\tau_{\rm rel} = \frac{\epsilon}{\sigma} = \frac{8.85\times 10^{-12}}{5.7\times 10^{7}} = 1.6\times 10^{-19}\ \text{sec}$$

By contrast, $\tau_{\rm rel}$ is of the order of days in a good dielectric. For good conductors, the above argument is not quite correct because it is based on the steady-state version of Ohm's law, $\boldsymbol{J} = \sigma\boldsymbol{E}$, which must be modified to take into account the transient dynamics of the conduction charges.

It turns out that the relaxation time $\tau_{\rm rel}$ is of the order of the collision time, which is typically 10^{-14} sec. We discuss this further in Sec. 1.9. See also Refs. [115–118].

1.7 Energy Flux and Energy Conservation

Because energy can be converted into different forms, the corresponding conservation equation (1.6.1) should have a non-zero term in the right-hand side corresponding to the rate by which energy is being lost from the fields into other forms, such as heat. Thus, we expect Eq. (1.6.1) to have the form:

$$\frac{\partial\rho_{\rm en}}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{J}_{\rm en} = \text{rate of energy loss} \tag{1.7.1}$$

Assuming the ordinary constitutive relations $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$, the quantities $\rho_{\rm en}, \boldsymbol{J}_{\rm en}$ describing the energy density and energy flux of the fields are defined as follows, where we introduce a change in notation:

$$
\begin{aligned}
\rho_{\mathrm{en}} &= w = \frac{1}{2}\epsilon|\boldsymbol{E}|^2 + \frac{1}{2}\mu|\boldsymbol{H}|^2 = \text{energy per unit volume} \\
\boldsymbol{J}_{\mathrm{en}} &= \boldsymbol{\mathcal{P}} = \boldsymbol{E}\times\boldsymbol{H} = \text{energy flux or Poynting vector}
\end{aligned}
\qquad (1.7.2)
$$

where $|\boldsymbol{E}|^2 = \boldsymbol{E}\cdot\boldsymbol{E}$. The quantities w and $\boldsymbol{\mathcal{P}}$ are measured in units of [joule/m^3] and [watt/m^2]. Using the identity $\nabla\cdot(\boldsymbol{E}\times\boldsymbol{H}) = \boldsymbol{H}\cdot\nabla\times\boldsymbol{E} - \boldsymbol{E}\cdot\nabla\times\boldsymbol{H}$, we find:

$$
\begin{aligned}
\frac{\partial w}{\partial t} + \nabla\cdot\boldsymbol{\mathcal{P}} &= \epsilon\frac{\partial \boldsymbol{E}}{\partial t}\cdot\boldsymbol{E} + \mu\frac{\partial \boldsymbol{H}}{\partial t}\cdot\boldsymbol{H} + \nabla\cdot(\boldsymbol{E}\times\boldsymbol{H}) \\
&= \frac{\partial \boldsymbol{D}}{\partial t}\cdot\boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t}\cdot\boldsymbol{H} + \boldsymbol{H}\cdot\nabla\times\boldsymbol{E} - \boldsymbol{E}\cdot\nabla\times\boldsymbol{H} \\
&= \left(\frac{\partial \boldsymbol{D}}{\partial t} - \nabla\times\boldsymbol{H}\right)\cdot\boldsymbol{E} + \left(\frac{\partial \boldsymbol{B}}{\partial t} + \nabla\times\boldsymbol{E}\right)\cdot\boldsymbol{H}
\end{aligned}
$$

Using Ampère's and Faraday's laws, the right-hand side becomes:

$$
\boxed{\frac{\partial w}{\partial t} + \nabla\cdot\boldsymbol{\mathcal{P}} = -\boldsymbol{J}\cdot\boldsymbol{E}} \qquad \text{(energy conservation)} \qquad (1.7.3)
$$

As we discussed in Eq. (1.2.6), the quantity $\boldsymbol{J}\cdot\boldsymbol{E}$ represents the ohmic losses, that is, the *power per unit volume* lost into heat from the fields. The integrated form of Eq. (1.7.3) is as follows, relative to the volume and surface of Fig. 1.6.1:

$$
-\oint_S \boldsymbol{\mathcal{P}}\cdot d\boldsymbol{S} = \frac{d}{dt}\int_V w\,dV + \int_V \boldsymbol{J}\cdot\boldsymbol{E}\,dV \qquad (1.7.4)
$$

It states that the total power *entering* a volume V through the surface S goes partially into increasing the field energy stored inside V and partially is lost into heat.

Example 1.7.1: Energy concepts can be used to derive the usual circuit formulas for capacitance, inductance, and resistance. Consider, for example, an ordinary plate capacitor with plates of area A separated by a distance l, and filled with a dielectric ϵ. The voltage between the plates is related to the electric field between the plates via $V = El$.

The energy density of the electric field between the plates is $w = \epsilon E^2/2$. Multiplying this by the volume between the plates, $A\cdot l$, will give the total energy stored in the capacitor. Equating this to the circuit expression $CV^2/2$, will yield the capacitance C:

$$
W = \frac{1}{2}\epsilon E^2\cdot Al = \frac{1}{2}CV^2 = \frac{1}{2}CE^2l^2 \quad\Rightarrow\quad C = \epsilon\,\frac{A}{l}
$$

Next, consider a solenoid with n turns wound around a cylindrical iron core of length l, cross-sectional area A, and permeability μ. The current through the solenoid wire is related to the magnetic field in the core through Ampère's law $Hl = nI$. It follows that the stored magnetic energy in the solenoid will be:

$$
W = \frac{1}{2}\mu H^2\cdot Al = \frac{1}{2}LI^2 = \frac{1}{2}L\frac{H^2l^2}{n^2} \quad\Rightarrow\quad L = n^2\mu\,\frac{A}{l}
$$

Finally, consider a resistor of length l, cross-sectional area A, and conductivity σ. The voltage drop across the resistor is related to the electric field along it via $V = El$. The current is assumed to be uniformly distributed over the cross-section A and will have density $J = \sigma E$.

The power dissipated into heat per unit volume is $JE = \sigma E^2$. Multiplying this by the resistor volume Al and equating it to the circuit expression $V^2/R = RI^2$ will give:

$$
(J\cdot E)\,(Al) = \sigma E^2\,(Al) = \frac{V^2}{R} = \frac{E^2l^2}{R} \quad\Rightarrow\quad R = \frac{1}{\sigma}\frac{l}{A}
$$

The same circuit expressions can, of course, be derived more directly using $Q = CV$, the magnetic flux $\Phi = LI$, and $V = RI$. □

Conservation laws may also be derived for the momentum carried by electromagnetic fields [40,1116]. It can be shown (see Problem 1.6) that the *momentum per unit volume* carried by the fields is given by:

$$
\boxed{\boldsymbol{G} = \boldsymbol{D}\times\boldsymbol{B} = \frac{1}{c^2}\boldsymbol{E}\times\boldsymbol{H} = \frac{1}{c^2}\boldsymbol{\mathcal{P}}} \qquad \text{(momentum density)} \qquad (1.7.5)
$$

where we set $\boldsymbol{D} = \epsilon\boldsymbol{E}$, $\boldsymbol{B} = \mu\boldsymbol{H}$, and $c = 1/\sqrt{\epsilon\mu}$. The quantity $\boldsymbol{J}_{\mathrm{mom}} = c\boldsymbol{G} = \boldsymbol{\mathcal{P}}/c$ will represent momentum flux, or pressure, if the fields are incident on a surface.

1.8 Harmonic Time Dependence

Maxwell's equations simplify considerably in the case of harmonic time dependence. Through the inverse Fourier transform, general solutions of Maxwell's equation can be built as linear combinations of single-frequency solutions:[†]

$$
\boldsymbol{E}(\mathbf{r},t) = \int_{-\infty}^{\infty}\boldsymbol{E}(\mathbf{r},\omega)\,e^{j\omega t}\,\frac{d\omega}{2\pi} \qquad (1.8.1)
$$

Thus, we assume that all fields have a time dependence $e^{j\omega t}$:

$$
\boldsymbol{E}(\mathbf{r},t) = \boldsymbol{E}(\mathbf{r})\,e^{j\omega t}, \quad \boldsymbol{H}(\mathbf{r},t) = \boldsymbol{H}(\mathbf{r})\,e^{j\omega t}
$$

where the phasor amplitudes $\boldsymbol{E}(\mathbf{r}), \boldsymbol{H}(\mathbf{r})$ are complex-valued. Replacing time derivatives by $\partial_t \to j\omega$, we may rewrite Eq. (1.1.1) in the form:

$$
\boxed{
\begin{aligned}
\nabla\times\boldsymbol{E} &= -j\omega\boldsymbol{B} \\
\nabla\times\boldsymbol{H} &= \boldsymbol{J} + j\omega\boldsymbol{D} \\
\nabla\cdot\boldsymbol{D} &= \rho \\
\nabla\cdot\boldsymbol{B} &= 0
\end{aligned}
} \qquad \text{(Maxwell's equations)} \qquad (1.8.2)
$$

In this book, we will consider the solutions of Eqs. (1.8.2) in three different contexts: (a) uniform plane waves propagating in dielectrics, conductors, and birefringent media, (b) guided waves propagating in hollow waveguides, transmission lines, and optical fibers, and (c) propagating waves generated by antennas and apertures.

[†] The $e^{j\omega t}$ convention is used in the engineering literature, and $e^{-i\omega t}$ in the physics literature. One can pass from one convention to the other by making the formal substitution $j \to -i$ in all the equations.

Next, we review some conventions regarding phasors and time averages. A real-valued sinusoid has the complex phasor representation:

$$\mathcal{A}(t)=|A|\cos(\omega t+\theta)\quad\Leftrightarrow\quad A(t)=Ae^{j\omega t}\tag{1.8.3}$$

where $A=|A|e^{j\theta}$. Thus, we have $\mathcal{A}(t)=\text{Re}\big[A(t)\big]=\text{Re}\big[Ae^{j\omega t}\big]$. The time averages of the quantities $\mathcal{A}(t)$ and $A(t)$ over one period $T=2\pi/\omega$ are zero.

The time average of the *product* of two harmonic quantities $\mathcal{A}(t)=\text{Re}\big[Ae^{j\omega t}\big]$ and $\mathcal{B}(t)=\text{Re}\big[Be^{j\omega t}\big]$ with phasors A,B is given by (see Problem 1.4):

$$\overline{\mathcal{A}(t)\mathcal{B}(t)}=\frac{1}{T}\int_0^T\mathcal{A}(t)\mathcal{B}(t)\,dt=\frac{1}{2}\text{Re}[AB^*]\tag{1.8.4}$$

In particular, the *mean-square* value is given by:

$$\overline{\mathcal{A}^2(t)}=\frac{1}{T}\int_0^T\mathcal{A}^2(t)\,dt=\frac{1}{2}\text{Re}[AA^*]=\frac{1}{2}|A|^2\tag{1.8.5}$$

Some interesting time averages in electromagnetic wave problems are the time averages of the energy density, the Poynting vector (energy flux), and the ohmic power losses per unit volume. Using the definition (1.7.2) and the result (1.8.4), we have for these time averages:

$$\begin{aligned}
w&=\frac{1}{2}\text{Re}\left[\frac{1}{2}\epsilon\mathbf{E}\cdot\mathbf{E}^*+\frac{1}{2}\mu\mathbf{H}\cdot\mathbf{H}^*\right] &&\text{(energy density)}\\
\boldsymbol{\mathcal{P}}&=\frac{1}{2}\text{Re}[\mathbf{E}\times\mathbf{H}^*] &&\text{(Poynting vector)}\\
\frac{dP_{\text{loss}}}{dV}&=\frac{1}{2}\text{Re}[\mathbf{J}_{\text{tot}}\cdot\mathbf{E}^*] &&\text{(ohmic losses)}
\end{aligned}\tag{1.8.6}$$

where $\mathbf{J}_{\text{tot}}=\mathbf{J}+j\omega\mathbf{D}$ is the total current in the right-hand side of Ampère's law and accounts for both conducting and dielectric losses. The time-averaged version of Poynting's theorem is discussed in Problem 1.5.

The expression (1.8.6) for the energy density w was derived under the assumption that both ϵ and μ were constants independent of frequency. In a dispersive medium, ϵ,μ become functions of frequency. In frequency bands where $\epsilon(\omega),\mu(\omega)$ are essentially real-valued, that is, where the medium is lossless, it can be shown [130] that the time-averaged energy density generalizes to:

$$w=\frac{1}{2}\text{Re}\left[\frac{1}{2}\frac{d(\omega\epsilon)}{d\omega}\mathbf{E}\cdot\mathbf{E}^*+\frac{1}{2}\frac{d(\omega\mu)}{d\omega}\mathbf{H}\cdot\mathbf{H}^*\right]\qquad\text{(lossless case)}\tag{1.8.7}$$

The derivation of (1.8.7) is as follows. Starting with Maxwell's equations (1.1.1) and without assuming any particular constitutive relations, we obtain:

$$\nabla\cdot\mathbf{E}\times\mathbf{H}=-\mathbf{E}\cdot\dot{\mathbf{D}}-\mathbf{H}\cdot\dot{\mathbf{B}}-\mathbf{J}\cdot\mathbf{E}\tag{1.8.8}$$

As in Eq. (1.7.3), we would like to interpret the first two terms in the right-hand side as the time derivative of the energy density, that is,

$$\frac{dw}{dt}=\mathbf{E}\cdot\dot{\mathbf{D}}+\mathbf{H}\cdot\dot{\mathbf{B}}$$

Anticipating a phasor-like representation, we may assume complex-valued fields and derive also the following relationship from Maxwell's equations:

$$\nabla\cdot\frac{1}{2}\text{Re}[\mathbf{E}\times\mathbf{H}^*]=-\frac{1}{2}\text{Re}[\mathbf{E}^*\cdot\dot{\mathbf{D}}]-\frac{1}{2}\text{Re}[\mathbf{H}^*\cdot\dot{\mathbf{B}}]-\frac{1}{2}\text{Re}[\mathbf{J}^*\cdot\mathbf{E}]\tag{1.8.9}$$

from which we may identify a "time-averaged" version of dw/dt,

$$\frac{d\bar{w}}{dt}=\frac{1}{2}\text{Re}[\mathbf{E}^*\cdot\dot{\mathbf{D}}]+\frac{1}{2}\text{Re}[\mathbf{H}^*\cdot\dot{\mathbf{B}}]\tag{1.8.10}$$

In a dispersive dielectric, the constitutive relation between $\mathbf{D}$ and $\mathbf{E}$ can be written as follows in the time and frequency domains:[†]

$$\mathbf{D}(t)=\int_{-\infty}^{\infty}\epsilon(t-t')\mathbf{E}(t')dt'\quad\Leftrightarrow\quad\mathbf{D}(\omega)=\epsilon(\omega)\mathbf{E}(\omega)\tag{1.8.11}$$

where the Fourier transforms are defined by

$$\epsilon(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\epsilon(\omega)e^{j\omega t}d\omega\quad\Leftrightarrow\quad\epsilon(\omega)=\int_{-\infty}^{\infty}\epsilon(t)e^{-j\omega t}dt\tag{1.8.12}$$

The time-derivative of $\mathbf{D}(t)$ is then

$$\dot{\mathbf{D}}(t)=\int_{-\infty}^{\infty}\dot{\epsilon}(t-t')\mathbf{E}(t')dt'\tag{1.8.13}$$

where it follows from Eq. (1.8.12) that

$$\dot{\epsilon}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}j\omega\epsilon(\omega)e^{j\omega t}d\omega\tag{1.8.14}$$

Following [130], we assume a quasi-harmonic representation for the electric field, $\mathbf{E}(t)=\mathbf{E}_0(t)e^{j\omega_0 t}$, where $\mathbf{E}_0(t)$ is a slowly-varying function of time. Equivalently, in the frequency domain we have $\mathbf{E}(\omega)=\mathbf{E}_0(\omega-\omega_0)$, assumed to be concentrated in a small neighborhood of ω_0, say, $|\omega-\omega_0|\le\Delta\omega$. Because $\epsilon(\omega)$ multiplies the narrowband function $\mathbf{E}(\omega)$, we may expand $\omega\epsilon(\omega)$ in a Taylor series around ω_0 and keep only the linear terms, that is, inside the integral (1.8.14), we may replace:

$$\omega\epsilon(\omega)=a_0+b_0(\omega-\omega_0)\,,\quad a_0=\omega_0\epsilon(\omega_0)\,,\quad b_0=\left.\frac{d[\omega\epsilon(\omega)]}{d\omega}\right|_{\omega_0}\tag{1.8.15}$$

Inserting this into Eq. (1.8.14), we obtain the approximation

$$\dot{\epsilon}(t)\simeq\frac{1}{2\pi}\int_{-\infty}^{\infty}[ja_0+b_0(j\omega-j\omega_0)]e^{j\omega t}d\omega=ja_0\delta(t)+b_0(\partial_t-j\omega_0)\delta(t)\tag{1.8.16}$$

where $\delta(t)$ the Dirac delta function. This approximation is justified only insofar as it is used inside Eq. (1.8.13). Inserting (1.8.16) into Eq. (1.8.13), we find

$$\begin{aligned}
\dot{\mathbf{D}}(t)&=\int_{-\infty}^{\infty}\big[ja_0\delta(t-t')+b_0(\partial_t-j\omega_0)\delta(t-t')\big]\mathbf{E}(t')dt'=\\
&=ja_0\mathbf{E}(t)+b_0(\partial_t-j\omega_0)\mathbf{E}(t)\\
&=ja_0\mathbf{E}_0(t)e^{j\omega_0 t}+b_0(\partial_t-j\omega_0)\big(\mathbf{E}_0(t)e^{j\omega_0 t}\big)\\
&=\big[ja_0\mathbf{E}_0(t)+b_0\dot{\mathbf{E}}_0(t)\big]e^{j\omega_0 t}
\end{aligned}\tag{1.8.17}$$

[†] To unclutter the notation, we are suppressing the dependence on the space coordinates $\mathbf{r}$.

Because we assume that $\epsilon(\omega)$ is real (i.e., lossless) in the vicinity of ω_0, it follows that:

$$\frac{1}{2}\,\mathrm{Re}\big[\mathbf{E}^*\cdot\dot{\mathbf{D}}\big] = \frac{1}{2}\,\mathrm{Re}\big[\mathbf{E}_0(t)^*\cdot\big(ja_0\mathbf{E}_0(t)+b_0\dot{\mathbf{E}}_0(t)\big)\big] = \frac{1}{2}b_0\,\mathrm{Re}\big[\mathbf{E}_0(t)^*\cdot\dot{\mathbf{E}}_0(t)\big]\,,\quad\text{or,}$$

$$\frac{1}{2}\,\mathrm{Re}\big[\mathbf{E}^*\cdot\dot{\mathbf{D}}\big] = \frac{d}{dt}\left[\frac{1}{4}b_0|\mathbf{E}_0(t)|^2\right] = \frac{d}{dt}\left[\frac{1}{4}\frac{d[\omega\epsilon(\omega)]_0}{d\omega}|\mathbf{E}_0(t)|^2\right] \tag{1.8.18}$$

Dropping the subscript 0, we see that the quantity under the time derivative in the right-hand side may be interpreted as a time-averaged energy density for the electric field. A similar argument can be given for the magnetic energy term of Eq. (1.8.7).

We will see in the next section that the energy density (1.8.7) consists of two parts: one part is the same as that in the vacuum case; the other part arises from the kinetic and potential energy stored in the polarizable molecules of the dielectric medium.

When Eq. (1.8.7) is applied to a plane wave propagating in a dielectric medium, one can show that (in the lossless case) the energy velocity coincides with the group velocity.

The generalization of these results to the case of a lossy medium has been studied extensively [130–144]. Eq. (1.8.7) has also been applied to the case of a "left-handed" medium in which both $\epsilon(\omega)$ and $\mu(\omega)$ are negative over certain frequency ranges. As argued by Veselago [353], such media must necessarily be dispersive in order to make Eq. (1.8.7) a positive quantity even though individually ϵ and μ are negative.

Analogous expressions to (1.8.7) may also be derived for the momentum density of a wave in a dispersive medium. In vacuum, the time-averaged momentum density is given by Eq. (1.7.5), that is,

$$\bar{\mathbf{G}} = \frac{1}{2}\,\mathrm{Re}\big[\epsilon_0\mu_0\,\mathbf{E}\times\mathbf{H}^*\big]$$

For the dispersive (and lossless) case this generalizes to [353,429]

$$\bar{\mathbf{G}} = \frac{1}{2}\,\mathrm{Re}\left[\epsilon\mu\mathbf{E}\times\mathbf{H}^* + \frac{\mathbf{k}}{2}\left(\frac{d\epsilon}{d\omega}|\mathbf{E}|^2 + \frac{d\mu}{d\omega}|\mathbf{H}|^2\right)\right] \tag{1.8.19}$$

1.9 Simple Models of Dielectrics, Conductors, and Plasmas

A simple model for the dielectric properties of a material is obtained by considering the motion of a bound electron in the presence of an applied electric field. As the electric field tries to separate the electron from the positively charged nucleus, it creates an electric dipole moment. Averaging this dipole moment over the volume of the material gives rise to a macroscopic dipole moment per unit volume.

A simple model for the dynamics of the displacement x of the bound electron is as follows (with $\dot{x} = dx/dt$):

$$m\ddot{x} = eE - kx - m\gamma\dot{x} \tag{1.9.1}$$

where we assumed that the electric field is acting in the x-direction and that there is a spring-like restoring force due to the binding of the electron to the nucleus, and a friction-type force proportional to the velocity of the electron.

The spring constant k is related to the resonance frequency of the spring via the relationship $\omega_0 = \sqrt{k/m}$, or, $k = m\omega_0^2$. Therefore, we may rewrite Eq. (1.9.1) as

$$\boxed{\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \frac{e}{m}E} \tag{1.9.2}$$

The limit $\omega_0 = 0$ corresponds to unbound electrons and describes the case of good conductors. The frictional term $\gamma\dot{x}$ arises from collisions that tend to slow down the electron. The parameter γ is a measure of the rate of collisions per unit time, and therefore, $\tau = 1/\gamma$ will represent the mean-time between collisions.

In a typical conductor, τ is of the order of 10^{-14} seconds, for example, for copper, $\tau = 2.4\times10^{-14}$ sec and $\gamma = 4.1\times10^{13}$ sec^{-1}. The case of a tenuous, collisionless, plasma can be obtained in the limit $\gamma = 0$. Thus, the above simple model can describe the following cases:

a. Dielectrics, $\omega_0 \neq 0, \gamma \neq 0$.
b. Conductors, $\omega_0 = 0, \gamma \neq 0$.
c. Collisionless Plasmas, $\omega_0 = 0, \gamma = 0$.

The basic idea of this model is that the applied electric field tends to separate positive from negative charges, thus, creating an electric dipole moment. In this sense, the model contains the basic features of other types of polarization in materials, such as ionic/molecular polarization arising from the separation of positive and negative ions by the applied field, or polar materials that have a permanent dipole moment.

Dielectrics

The applied electric field $E(t)$ in Eq. (1.9.2) can have any time dependence. In particular, if we assume it is sinusoidal with frequency ω, $E(t) = Ee^{j\omega t}$, then, Eq. (1.9.2) will have the solution $x(t) = xe^{j\omega t}$, where the phasor x must satisfy:

$$-\omega^2 x + j\omega\gamma x + \omega_0^2 x = \frac{e}{m}E$$

which is obtained by replacing time derivatives by $\partial_t \to j\omega$. Its solution is:

$$x = \frac{\dfrac{e}{m}E}{\omega_0^2 - \omega^2 + j\omega\gamma} \tag{1.9.3}$$

The corresponding velocity of the electron will also be sinusoidal $v(t) = ve^{j\omega t}$, where $v = \dot{x} = j\omega x$. Thus, we have:

$$v = j\omega x = \frac{j\omega\dfrac{e}{m}E}{\omega_0^2 - \omega^2 + j\omega\gamma} \tag{1.9.4}$$

From Eqs. (1.9.3) and (1.9.4), we can find the polarization per unit volume P. We assume that there are N such elementary dipoles per unit volume. The individual electric dipole moment is $p = ex$. Therefore, the polarization per unit volume will be:

$$P = Np = Nex = \frac{\dfrac{Ne^2}{m}E}{\omega_0^2 - \omega^2 + j\omega\gamma} \equiv \epsilon_0\chi(\omega)E \tag{1.9.5}$$

The electric flux density will be then:

$$D = \epsilon_0 E + P = \epsilon_0\big(1 + \chi(\omega)\big)E \equiv \epsilon(\omega)E$$

where the effective permittivity $\epsilon(\omega)$ is:

$$\epsilon(\omega)=\epsilon_0+\frac{\dfrac{Ne^2}{m}}{\omega_0^2-\omega^2+j\omega\gamma} \tag{1.9.6}$$

This can be written in a more convenient form, as follows:

$$\boxed{\epsilon(\omega)=\epsilon_0+\frac{\epsilon_0\omega_p^2}{\omega_0^2-\omega^2+j\omega\gamma}} \tag{1.9.7}$$

where ω_p^2 is the so-called *plasma frequency* of the material defined by:

$$\boxed{\omega_p^2=\frac{Ne^2}{\epsilon_0 m}}\qquad\text{(plasma frequency)} \tag{1.9.8}$$

The model defined by (1.9.7) is know as a "Lorentz dielectric." The corresponding susceptibility, defined through $\epsilon(\omega)=\epsilon_0\big(1+\chi(\omega)\big)$, is:

$$\chi(\omega)=\frac{\omega_p^2}{\omega_0^2-\omega^2+j\omega\gamma} \tag{1.9.9}$$

For a dielectric, we may assume $\omega_0\neq 0$. Then, the low-frequency limit ($\omega=0$) of Eq. (1.9.7), gives the nominal dielectric constant:

$$\epsilon(0)=\epsilon_0+\epsilon_0\frac{\omega_p^2}{\omega_0^2}=\epsilon_0+\frac{Ne^2}{m\omega_0^2} \tag{1.9.10}$$

The real and imaginary parts of $\epsilon(\omega)$ characterize the *refractive* and *absorptive* properties of the material. By convention, we define the imaginary part with the negative sign (because we use $e^{j\omega t}$ time dependence):

$$\epsilon(\omega)=\epsilon'(\omega)-j\epsilon''(\omega) \tag{1.9.11}$$

It follows from Eq. (1.9.7) that:

$$\epsilon'(\omega)=\epsilon_0+\frac{\epsilon_0\omega_p^2(\omega_0^2-\omega^2)}{(\omega^2-\omega_0^2)^2+\gamma^2\omega^2}\,,\qquad \epsilon''(\omega)=\frac{\epsilon_0\omega_p^2\omega\gamma}{(\omega^2-\omega_0^2)^2+\gamma^2\omega^2} \tag{1.9.12}$$

Fig. 1.9.1 shows a plot of $\epsilon'(\omega)$ and $\epsilon''(\omega)$. Around the resonant frequency ω_0, the real part $\epsilon'(\omega)$ behaves in an anomalous manner, that is, it drops rapidly with frequency to values less than ϵ_0 and the material exhibits strong absorption. The term "normal dispersion" refers to an $\epsilon'(\omega)$ that is an increasing function of ω, as is the case to the far left and right of the resonant frequency.

Real dielectric materials exhibit, of course, several such resonant frequencies corresponding to various vibrational modes and polarization mechanisms (e.g., electronic, ionic, etc.) The permittivity becomes the sum of such terms:

$$\epsilon(\omega)=\epsilon_0+\epsilon_0\sum_i\frac{N_ie_i^2/m_i\epsilon_0}{\omega_i^2-\omega^2+j\omega\gamma_i} \tag{1.9.13}$$

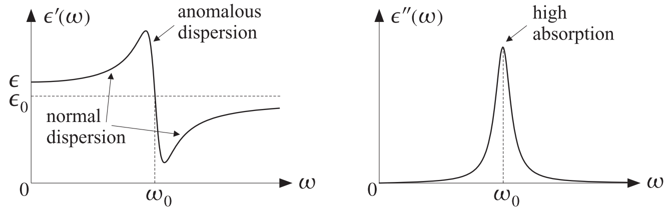


Fig. 1.9.1 Real and imaginary parts of the effective permittivity $\epsilon(\omega)$.

A more correct quantum-mechanical treatment leads essentially to the same formula:

$$\epsilon(\omega)=\epsilon_0+\epsilon_0\sum_{j>i}\frac{f_{ji}(N_i-N_j)e^2/m\epsilon_0}{\omega_{ji}^2-\omega^2+j\omega\gamma_{ji}} \tag{1.9.14}$$

where ω_{ji} are transition frequencies between energy levels, that is, $\omega_{ji}=(E_j-E_i)/\hbar$, and N_i, N_j are the populations of the lower, E_i, and upper, E_j, energy levels. The quantities f_{ji} are called "oscillator strengths." For example, for a two-level atom we have:

$$\epsilon(\omega)=\epsilon_0+\epsilon_0\frac{f\omega_p^2}{\omega_0^2-\omega^2+j\omega\gamma} \tag{1.9.15}$$

where we defined:

$$\omega_0=\omega_{21}\,,\quad f=f_{21}\frac{N_1-N_2}{N_1+N_2}\,,\quad \omega_p^2=\frac{(N_1+N_2)e^2}{m\epsilon_0}$$

Normally, lower energy states are more populated, $N_i>N_j$, and the material behaves as a classical absorbing dielectric. However, if there is population inversion, $N_i<N_j$, then the corresponding permittivity term changes sign. This leads to a *negative* imaginary part, $\epsilon''(\omega)$, representing a *gain*. Fig. 1.9.2 shows the real and imaginary parts of Eq. (1.9.15) for the case of a negative effective oscillator strength $f=-1$.

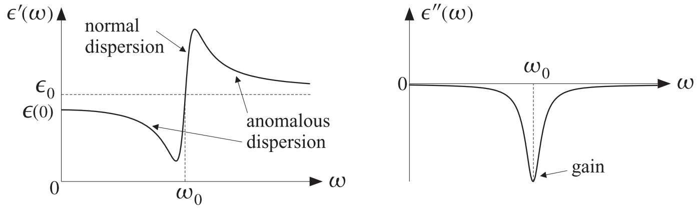


Fig. 1.9.2 Effective permittivity in a two-level gain medium with $f=-1$.

The normal and anomalous dispersion bands still correspond to the bands where the real part $\epsilon'(\omega)$ is an increasing or decreasing, respectively, function of frequency. But now the normal behavior is only in the neighborhood of the resonant frequency, whereas far from it, the behavior is anomalous.

Setting $n(\omega)=\sqrt{\epsilon(\omega)/\epsilon_0}$ for the refractive index, Eq. (1.9.13) can be written in the following form, known as the Sellmeier equation:

$$n^2(\omega)=1+\sum_i\frac{B_i}{\omega_i^2-\omega^2+j\omega\gamma_i}\tag{1.9.16}$$

In practice, Eq. (1.9.16) is applied in frequency ranges that are far from any resonance so that one can effectively set $\gamma_i=0$:

$$n^2(\omega)=1+\sum_i\frac{B_i}{\omega_i^2-\omega^2}=1+\sum_i\frac{B_i\lambda^2}{\lambda^2-\lambda_i^2}\qquad\text{(Sellmeier equation)}\tag{1.9.17}$$

where λ,λ_i denote the corresponding free-space wavelengths (e.g., $\lambda=2\pi c/\omega$). In practice, refractive index data are fitted to Eq. (1.9.17) using 2–4 terms over a desired frequency range. For example, fused silica (SiO_2) is very accurately represented over the range $0.2\le\lambda\le3.7$ μm by the formula [124], where λ and λ_i are in units of μm:

$$n^2=1+\frac{0.6961663\,\lambda^2}{\lambda^2-(0.0684043)^2}+\frac{0.4079426\,\lambda^2}{\lambda^2-(0.1162414)^2}+\frac{0.8974794\,\lambda^2}{\lambda^2-(9.896161)^2}\tag{1.9.18}$$

Conductors

The conductivity properties of a material are described by Ohm's law, Eq. (1.3.15). To derive this law from our simple model, we use the relationship $J=\rho v$, where the volume density of the conduction charges is $\rho=Ne$. It follows from Eq. (1.9.4) that

$$J=\rho v=Nev=\frac{j\omega\dfrac{Ne^2}{m}E}{\omega_0^2-\omega^2+j\omega\gamma}\equiv\sigma(\omega)E$$

and therefore, we identify the conductivity $\sigma(\omega)$:

$$\sigma(\omega)=\frac{j\omega\dfrac{Ne^2}{m}}{\omega_0^2-\omega^2+j\omega\gamma}=\frac{j\omega\epsilon_0\omega_p^2}{\omega_0^2-\omega^2+j\omega\gamma}\tag{1.9.19}$$

We note that $\sigma(\omega)/j\omega$ is essentially the electric susceptibility considered above. Indeed, we have $J=Nev=Nej\omega x=j\omega P$, and thus, $P=J/j\omega=(\sigma(\omega)/j\omega)E$. It follows that $\epsilon(\omega)-\epsilon_0=\sigma(\omega)/j\omega$, and

$$\epsilon(\omega)=\epsilon_0+\frac{\epsilon_0\omega_p^2}{\omega_0^2-\omega^2+j\omega\gamma}=\epsilon_0+\frac{\sigma(\omega)}{j\omega}\tag{1.9.20}$$

Since in a metal the conduction charges are unbound, we may take $\omega_0=0$ in Eq. (1.9.19). After canceling a common factor of $j\omega$, we obtain:

$$\boxed{\sigma(\omega)=\frac{\epsilon_0\omega_p^2}{\gamma+j\omega}}\tag{1.9.21}$$

The model defined by (1.9.21) is know as the "Drude model." The nominal conductivity is obtained at the low-frequency limit, $\omega=0$:

$$\boxed{\sigma=\frac{\epsilon_0\omega_p^2}{\gamma}=\frac{Ne^2}{m\gamma}}\qquad\text{(nominal conductivity)}\tag{1.9.22}$$

Example 1.9.1: Copper has a mass density of 8.9×10^6 gr/m^3 and atomic weight of 63.54 (grams per mole.) Using Avogadro's number of 6×10^{23} atoms per mole, and assuming one conduction electron per atom, we find for the volume density N:

$$N=\frac{6\times10^{23}\,\dfrac{\text{atoms}}{\text{mole}}}{63.54\,\dfrac{\text{gr}}{\text{mole}}}\Big(8.9\times10^6\,\frac{\text{gr}}{\text{m}^3}\Big)\Big(1\,\frac{\text{electron}}{\text{atom}}\Big)=8.4\times10^{28}\ \text{electrons/m}^3$$

It follows that:

$$\sigma=\frac{Ne^2}{m\gamma}=\frac{(8.4\times10^{28})(1.6\times10^{-19})^2}{(9.1\times10^{-31})(4.1\times10^{13})}=5.8\times10^7\ \text{Siemens/m}$$

where we used $e=1.6\times10^{-19}$, $m=9.1\times10^{-31}$, $\gamma=4.1\times10^{13}$. The plasma frequency of copper can be calculated by

$$f_p=\frac{\omega_p}{2\pi}=\frac{1}{2\pi}\sqrt{\frac{Ne^2}{m\epsilon_0}}=2.6\times10^{15}\ \text{Hz}$$

which lies in the ultraviolet range. For frequencies such that $\omega\ll\gamma$, the conductivity (1.9.21) may be considered to be independent of frequency and equal to the dc value of Eq. (1.9.22). This frequency range covers most present-day RF applications. For example, assuming $\omega\le0.1\gamma$, we find $f\le0.1\gamma/2\pi=653$ GHz. ☐

So far, we assumed sinusoidal time dependence and worked with the steady-state responses. Next, we discuss the transient dynamical response of a conductor subject to an arbitrary time-varying electric field $E(t)$.

Ohm's law can be expressed either in the frequency-domain or in the time-domain with the help the Fourier transform pair of equations:

$$J(\omega)=\sigma(\omega)E(\omega)\qquad\Leftrightarrow\qquad J(t)=\int_{-\infty}^{t}\sigma(t-t')E(t')dt'\tag{1.9.23}$$

where $\sigma(t)$ is the causal inverse Fourier transform of $\sigma(\omega)$. For the simple model of Eq. (1.9.21), we have:

$$\sigma(t)=\epsilon_0\omega_p^2e^{-\gamma t}u(t)\tag{1.9.24}$$

where $u(t)$ is the unit-step function. As an example, suppose the electric field $E(t)$ is a constant electric field that is suddenly turned on at $t=0$, that is, $E(t)=Eu(t)$. Then, the time response of the current will be:

$$J(t)=\int_0^t\epsilon_0\omega_p^2e^{-\gamma(t-t')}E\,dt'=\frac{\epsilon_0\omega_p^2}{\gamma}E\big(1-e^{-\gamma t}\big)=\sigma E\big(1-e^{-\gamma t}\big)$$

where $\sigma = \epsilon_0\omega_p^2/\gamma$ is the nominal conductivity of the material.

Thus, the current starts out at zero and builds up to the steady-state value of $J = \sigma E$, which is the conventional form of Ohm's law. The rise time constant is $\tau = 1/\gamma$. We saw above that τ is extremely small—of the order of 10^{-14} sec—for good conductors.

The building up of the current can also be understood in terms of the equation of motion of the conducting charges. Writing Eq. (1.9.2) in terms of the velocity of the charge, we have:

$$\dot{v}(t)+\gamma v(t)=\frac{e}{m}E(t)$$

Assuming $E(t)=Eu(t)$, we obtain the convolutional solution:

$$v(t)=\int_0^t e^{-\gamma(t-t')}\frac{e}{m}E(t')dt' = \frac{e}{m\gamma}E(1-e^{-\gamma t})$$

For large t, the velocity reaches the steady-state value $v_\infty = (e/m\gamma)E$, which reflects the balance between the accelerating electric field force and the retarding frictional force, that is, $m\gamma v_\infty = eE$. The quantity $e/m\gamma$ is called the *mobility* of the conduction charges. The steady-state current density results in the conventional Ohm's law:

$$J = Nev_\infty = \frac{Ne^2}{m\gamma}E = \sigma E$$

Charge Relaxation in Conductors

Next, we discuss the issue of charge relaxation in good conductors [115–118]. Writing (1.9.23) three-dimensionally and using (1.9.24), Ohm's law reads in the time domain:

$$\boldsymbol{J}(\mathbf{r},t)=\omega_p^2\int_{-\infty}^t e^{-\gamma(t-t')}\epsilon_0\,\boldsymbol{E}(\mathbf{r},t')\,dt' \tag{1.9.25}$$

Taking the divergence of both sides and using charge conservation, $\nabla\cdot\boldsymbol{J}+\dot{\rho}=0$, and Gauss's law, $\epsilon_0\nabla\cdot\boldsymbol{E}=\rho$, we obtain the following integro-differential equation for the charge density $\rho(\mathbf{r},t)$:

$$-\dot{\rho}(\mathbf{r},t)=\nabla\cdot\boldsymbol{J}(\mathbf{r},t)=\omega_p^2\int_{-\infty}^t e^{-\gamma(t-t')}\epsilon_0\nabla\cdot\boldsymbol{E}(\mathbf{r},t')dt' = \omega_p^2\int_{-\infty}^t e^{-\gamma(t-t')}\rho(\mathbf{r},t')dt'$$

Differentiating both sides with respect to t, we find that ρ satisfies the second-order differential equation:

$$\ddot{\rho}(\mathbf{r},t)+\gamma\dot{\rho}(\mathbf{r},t)+\omega_p^2\rho(\mathbf{r},t)=0 \tag{1.9.26}$$

whose solution is easily verified to be a linear combination of:

$$e^{-\gamma t/2}\cos(\omega_{\rm relax}t)\,,\quad e^{-\gamma t/2}\sin(\omega_{\rm relax}t)\,,\quad\text{where}\quad \omega_{\rm relax}=\sqrt{\omega_p^2-\frac{\gamma^2}{4}}$$

Thus, the charge density is an exponentially decaying sinusoid with a relaxation time constant that is twice the collision time $\tau = 1/\gamma$:

$$\boxed{\tau_{\rm relax}=\frac{2}{\gamma}=2\tau}\qquad\text{(relaxation time constant)} \tag{1.9.27}$$

Typically, $\omega_p \gg \gamma$, so that $\omega_{\rm relax}$ is practically equal to ω_p. For example, using the numerical data of Example 1.9.1, we find for copper $\tau_{\rm relax} = 2\tau = 5{\times}10^{-14}$ sec. We calculate also: $f_{\rm relax} = \omega_{\rm relax}/2\pi = 2.6{\times}10^{15}$ Hz. In the limit $\gamma\to\infty$, or $\tau\to 0$, Eq. (1.9.26) reduces to the naive relaxation equation (1.6.3) (see Problem 1.9).

In addition to charge relaxation, the total relaxation time depends on the time it takes for the electric and magnetic fields to be extinguished from the inside of the conductor, as well as the time it takes for the accumulated surface charge densities to settle, the motion of the surface charges being damped because of ohmic losses. Both of these times depend on the geometry and size of the conductor [117].

Power Losses

To describe a material with *both* dielectric and conductivity properties, we may take the susceptibility to be the sum of two terms, one describing bound polarized charges and the other unbound conduction charges. Assuming different parameters $\{\omega_0, \omega_p, \gamma\}$ for each term, we obtain the total permittivity:

$$\epsilon(\omega)=\epsilon_0+\frac{\epsilon_0\omega_{dp}^2}{\omega_{d0}^2-\omega^2+j\omega\gamma_d}+\frac{\epsilon_0\omega_{cp}^2}{j\omega(\gamma_c+j\omega)} \tag{1.9.28}$$

Denoting the first two terms by $\epsilon_d(\omega)$ and the third by $\sigma_c(\omega)/j\omega$, we obtain the total effective permittivity of such a material:

$$\boxed{\epsilon(\omega)=\epsilon_d(\omega)+\frac{\sigma_c(\omega)}{j\omega}}\qquad\text{(effective permittivity)} \tag{1.9.29}$$

In the low-frequency limit, $\omega = 0$, the quantities $\epsilon_d(0)$ and $\sigma_c(0)$ represent the nominal dielectric constant and conductivity of the material. We note also that we can write Eq. (1.9.29) in the form:

$$j\omega\epsilon(\omega)=\sigma_c(\omega)+j\omega\epsilon_d(\omega) \tag{1.9.30}$$

These two terms characterize the relative importance of the *conduction* current and the *displacement* (polarization) current. The right-hand side in Ampère's law gives the total effective current:

$$J_{\rm tot}=J+\frac{\partial D}{\partial t}=J+j\omega D=\sigma_c(\omega)E+j\omega\epsilon_d(\omega)E=j\omega\epsilon(\omega)E$$

where the term $J_{\rm disp} = \partial D/\partial t = j\omega\epsilon_d(\omega)E$ represents the displacement current. The relative strength between conduction and displacement currents is the ratio:

$$\left|\frac{J_{\rm cond}}{J_{\rm disp}}\right|=\frac{|\sigma_c(\omega)E|}{|j\omega\epsilon_d(\omega)E|}=\frac{|\sigma_c(\omega)|}{|\omega\epsilon_d(\omega)|} \tag{1.9.31}$$

This ratio is frequency-dependent and establishes a *dividing line* between a good conductor and a good dielectric. If the ratio is much larger than unity (typically, greater than 10), the material behaves as a good conductor at that frequency; if the ratio is much smaller than one (typically, less than 0.1), then the material behaves as a good dielectric.

Example 1.9.2: This ratio can take a very wide range of values. For example, assuming a frequency of 1 GHz and using (for illustration purposes) the dc-values of the dielectric constants and conductivities, we find:

$$\left|\frac{J_{\text{cond}}}{J_{\text{disp}}}\right| = \frac{\sigma}{\omega\epsilon} = \begin{cases} 10^9 & \text{for copper with } \sigma = 5.8\times10^7 \text{ S/m and } \epsilon = \epsilon_0 \\ 1 & \text{for seawater with } \sigma = 4 \text{ S/m and } \epsilon = 72\epsilon_0 \\ 10^{-9} & \text{for a glass with } \sigma = 10^{-10} \text{ S/m and } \epsilon = 2\epsilon_0 \end{cases}$$

Thus, the ratio varies over 18 orders of magnitude! If the frequency is reduced by a factor of ten to 100 MHz, then all the ratios get multiplied by 10. In this case, seawater acts like a good conductor. □

The time-averaged ohmic power losses per unit volume within a lossy material are given by Eq. (1.8.6). Writing $\epsilon(\omega) = \epsilon'(\omega) - j\epsilon''(\omega)$, we have:

$$\boldsymbol{J}_{\text{tot}} = j\omega\epsilon(\omega)\boldsymbol{E} = j\omega\epsilon'(\omega)\boldsymbol{E} + \omega\epsilon''(\omega)\boldsymbol{E}$$

Denoting $|\boldsymbol{E}|^2 = \boldsymbol{E}\cdot\boldsymbol{E}^*$, it follows that:

$$\frac{dP_{\text{loss}}}{dV} = \frac{1}{2}\text{Re}\big[\boldsymbol{J}_{\text{tot}}\cdot\boldsymbol{E}^*\big] = \frac{1}{2}\omega\epsilon''(\omega)\,|\boldsymbol{E}|^2 \qquad \text{(ohmic losses)} \tag{1.9.32}$$

Writing $\epsilon_d(\omega) = \epsilon'_d(\omega) - j\epsilon''_d(\omega)$ and assuming that the conductivity $\sigma_c(\omega)$ is real-valued for the frequency range of interest (as was discussed in Example 1.9.1), we find by equating real and imaginary parts of Eq. (1.9.29):

$$\epsilon'(\omega) = \epsilon'_d(\omega)\,, \qquad \epsilon''(\omega) = \epsilon''_d(\omega) + \frac{\sigma_c(\omega)}{\omega} \tag{1.9.33}$$

Then, the power losses can be written in a form that separates the losses due to conduction and those due to the polarization properties of the dielectric:

$$\frac{dP_{\text{loss}}}{dV} = \frac{1}{2}\big(\sigma_c(\omega) + \omega\epsilon''_d(\omega)\big)\,|\boldsymbol{E}|^2 \qquad \text{(ohmic losses)} \tag{1.9.34}$$

A convenient way to quantify the losses is by means of the *loss tangent* defined in terms of the real and imaginary parts of the effective permittivity:

$$\boxed{\tan\theta = \frac{\epsilon''(\omega)}{\epsilon'(\omega)}} \qquad \text{(loss tangent)} \tag{1.9.35}$$

where θ is the *loss angle*. Eq. (1.9.35) may be written as the sum of two loss tangents, one due to conduction and one due to polarization. Using Eq. (1.9.33), we have:

$$\tan\theta = \frac{\sigma_c(\omega) + \omega\epsilon''_d(\omega)}{\omega\epsilon'_d(\omega)} = \frac{\sigma_c(\omega)}{\omega\epsilon'_d(\omega)} + \frac{\epsilon''_d(\omega)}{\epsilon'_d(\omega)} = \tan\theta_c + \tan\theta_d \tag{1.9.36}$$

The ohmic loss per unit volume can be expressed in terms of the loss tangent as:

$$\frac{dP_{\text{loss}}}{dV} = \frac{1}{2}\omega\epsilon'_d(\omega)\tan\theta\,|\boldsymbol{E}|^2 \qquad \text{(ohmic losses)} \tag{1.9.37}$$

Plasmas

To describe a collisionless plasma, such as the ionosphere, the simple model considered in the previous sections can be specialized by choosing $\omega_0 = \gamma = 0$. Thus, the conductivity given by Eq. (1.9.21) becomes pure imaginary:

$$\sigma(\omega) = \frac{\epsilon_0\omega_p^2}{j\omega}$$

The corresponding effective permittivity of Eq. (1.9.20) becomes purely real:

$$\boxed{\epsilon(\omega) = \epsilon_0 + \frac{\sigma(\omega)}{j\omega} = \epsilon_0\left(1 - \frac{\omega_p^2}{\omega^2}\right)} \tag{1.9.38}$$

The plasma frequency can be calculated from $\omega_p^2 = Ne^2/m\epsilon_0$. In the ionosphere the electron density is typically $N = 10^{12}$, which gives $f_p = 9$ MHz.

We will see in Sec. 2.6 that the propagation wavenumber of an electromagnetic wave propagating in a dielectric/conducting medium is given in terms of the effective permittivity by:

$$k = \omega\sqrt{\mu\epsilon(\omega)}$$

It follows that for a plasma:

$$k = \omega\sqrt{\mu_0\epsilon_0(1 - \omega_p^2/\omega^2)} = \frac{1}{c}\sqrt{\omega^2 - \omega_p^2} \tag{1.9.39}$$

where we used $c = 1/\sqrt{\mu_0\epsilon_0}$.

If $\omega > \omega_p$, the electromagnetic wave propagates without attenuation within the plasma. But if $\omega < \omega_p$, the wavenumber k becomes imaginary and the wave gets attenuated. At such frequencies, a wave incident (normally) on the ionosphere from the ground cannot penetrate and gets reflected back.

Energy Density in a Lossless Dispersive Dielectric

The lossless case is obtained from Eq. (1.9.7) by setting $\gamma = 0$, which is equivalent to assuming that ω is far from the resonance ω_0. In this case the permittivity is:

$$\epsilon(\omega) = \epsilon_0\left[1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2}\right]$$

from which it follows that:

$$\frac{d(\omega\epsilon)}{d\omega} = \epsilon_0\left[1 + \frac{\omega_p^2(\omega^2 + \omega_0^2)}{(\omega_0^2 - \omega^2)^2}\right] \tag{1.9.40}$$

Thus, the electric part of the energy density (1.8.7) will be:

$$\bar{w}_e = \frac{1}{4}\frac{d(\omega\epsilon)}{d\omega}|E|^2 = \frac{1}{4}\epsilon_0|E|^2\left[1 + \frac{\omega_p^2(\omega^2 + \omega_0^2)}{(\omega_0^2 - \omega^2)^2}\right] \tag{1.9.41}$$

This expression can be given a nice interpretation: The first term on the right is the energy density in vacuum and the second corresponds to the mechanical (kinetic and potential) energy of the polarization charges [131,154]. Indeed, the displacement x and velocity $v = \dot{x}$ of the polarization charges are in this case:

$$x = \frac{eE/m}{\omega_0^2 - \omega^2}\,, \quad v = j\omega x$$

The time-averaged mechanical energy (per unit volume) is obtained by adding the kinetic and potential energies:

$$\begin{aligned} \bar{w}_{\text{mech}} &= \frac{1}{2}\,\text{Re}\left[N\left(\frac{1}{2}m|v|^2 + \frac{1}{2}m\omega_0^2|x|^2\right)\right] = \frac{1}{4}Nm(\omega^2 + \omega_0^2)|x|^2 \\ &= \frac{1}{4}\,\frac{Nm(\omega^2+\omega_0^2)e^2|E|^2/m^2}{(\omega_0^2-\omega^2)^2} = \frac{1}{4}\epsilon_0|E|^2\left[\frac{\omega_p^2(\omega^2+\omega_0^2)}{(\omega_0^2-\omega^2)^2}\right] \end{aligned}$$

where we used the definition (1.9.8) of the plasma frequency. It follows that Eq. (1.9.41) can be written as the sum:

$$\bar{w}_e = \frac{1}{4}\frac{d(\omega\epsilon)}{d\omega}|E|^2 = \frac{1}{4}\epsilon_0|E|^2 + \bar{w}_{\text{mech}} = \bar{w}_{\text{vac}} + \bar{w}_{\text{mech}} \tag{1.9.42}$$

1.10 Kramers-Kronig Dispersion Relations

The convolutional form of Eq. (1.3.13) implies causality, that is, the value of $\boldsymbol{D}(\mathbf{r},t)$ at the present time t depends only on the past values of $\boldsymbol{E}(\mathbf{r},t')$, $t' \le t$.

This condition is equivalent to requiring that the dielectric response $\epsilon(t)$ be a right-sided (causal) function of time, that is, $\epsilon(t) = 0$ for $t < 0$. Then, Eq. (1.3.13) may be written as ordinary convolution by extending the integration range over all times:

$$\boldsymbol{D}(\mathbf{r},t) = \int_{-\infty}^{t}\epsilon(t-t')\boldsymbol{E}(\mathbf{r},t')\,dt' = \int_{-\infty}^{\infty}\epsilon(t-t')\boldsymbol{E}(\mathbf{r},t')\,dt'$$

Because $\boldsymbol{D}(\mathbf{r},t) = \epsilon_0\boldsymbol{E}(\mathbf{r},t) + \boldsymbol{P}(\mathbf{r},t)$, we may define the time-domain susceptibility function $\chi(t)$ through:

$$\epsilon(t) = \epsilon_0\delta(t) + \epsilon_0\chi(t) \tag{1.10.1}$$

where $\delta(t)$ is the Dirac delta function. Therefore, if $\epsilon(t)$ is causal, so is $\chi(t)$. The polarization is then given by:

$$\boldsymbol{P}(\mathbf{r},t) = \epsilon_0\int_{-\infty}^{t}\chi(t-t')\boldsymbol{E}(\mathbf{r},t')\,dt' = \epsilon_0\int_{-\infty}^{\infty}\chi(t-t')\boldsymbol{E}(\mathbf{r},t')\,dt' \tag{1.10.2}$$

In the frequency domain, this becomes multiplicative: $\boldsymbol{P}(\mathbf{r},\omega) = \epsilon_0\chi(\omega)\boldsymbol{E}(\mathbf{r},\omega)$. The Kramers-Kronig relations are the frequency-domain expression of causality and relate the real and imaginary parts of the susceptibility function $\chi(\omega)$. Here, the functions $\chi(t)$ and $\chi(\omega)$ are Fourier transform pairs:

$$\chi(\omega) = \int_{-\infty}^{\infty}\chi(t)e^{-j\omega t}\,dt \quad\Leftrightarrow\quad \chi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\chi(\omega)e^{j\omega t}\,d\omega \tag{1.10.3}$$

The causality condition, $\chi(t) = 0$ for $t < 0$, can be expressed in terms of the unit-step function $u(t)$ in the equivalent manner:

$$\chi(t) = \chi(t)u(t)\,, \quad \text{for all } t \tag{1.10.4}$$

Using the property that the Fourier transform of a product of two time functions is the convolution of their Fourier transforms, it follows that Eq. (1.10.4) can be written in the equivalent frequency-domain form:

$$\chi(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\chi(\omega')U(\omega-\omega')\,d\omega' \tag{1.10.5}$$

where $U(\omega)$ is the Fourier transform of the unit-step. Eq. (1.10.5) is essentially the Kramers-Kronig relation. The function $U(\omega)$ is given by the well-known expression:

$$U(\omega) = \lim_{\epsilon\to 0+}\frac{1}{j\omega+\epsilon} = \mathcal{P}\frac{1}{j\omega} + \pi\delta(\omega) \tag{1.10.6}$$

where $\mathcal{P}$ denotes the "principal value." Inserting (1.10.6) into (1.10.5), we have:

$$\begin{aligned} \chi(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\chi(\omega')\left[\mathcal{P}\frac{1}{j(\omega-\omega')} + \pi\delta(\omega-\omega')\right]d\omega' \\ &= \frac{1}{2\pi j}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi(\omega')}{\omega-\omega'}\,d\omega' + \frac{1}{2}\chi(\omega) \end{aligned}$$

Rearranging terms and canceling a factor of $1/2$, we obtain the Kramers-Kronig relation in its complex-valued form:†

$$\chi(\omega) = \frac{1}{\pi j}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi(\omega')}{\omega-\omega'}\,d\omega' \qquad \text{(Kramers-Kronig)} \tag{1.10.7}$$

The reason for applying this relation to $\chi(\omega)$ instead of $\epsilon(\omega)$ is that $\chi(\omega)$ falls off sufficiently fast for large ω to make the integral in (1.10.5) convergent, whereas $\epsilon(\omega)$ tends to the constant ϵ_0.

Setting $\chi(\omega) = \chi_r(\omega) - j\chi_i(\omega)$ and separating (1.10.7) into its real and imaginary parts, we obtain the conventional form of the Kramers-Kronig dispersion relations:

$$\boxed{\begin{aligned} \chi_r(\omega) &= \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi_i(\omega')}{\omega'-\omega}\,d\omega' \\ \chi_i(\omega) &= -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi_r(\omega')}{\omega'-\omega}\,d\omega' \end{aligned}} \qquad \text{(Kramers-Kronig relations)} \tag{1.10.8}$$

Because the time-response $\chi(t)$ is real-valued, its Fourier transform $\chi(\omega)$ will satisfy the Hermitian symmetry property $\chi(-\omega) = \chi^*(\omega)$, which is equivalent to the even symmetry of its real part, $\chi_r(-\omega) = \chi_r(\omega)$, and the odd symmetry of its imaginary part,

---

† The right-hand side (without the j) in (1.10.7) is known as a Hilbert transform. Exchanging the roles of t and ω, such transforms, known also as 90° phase shifters, are used widely in signal processing for generating single-sideband communications signals.

$\chi_i(-\omega)= -\chi_i(\omega)$. Taking advantage of these symmetries, the range of integration in (1.10.8) can be folded in half resulting in:

$$\begin{aligned}\chi_r(\omega) &= \frac{2}{\pi}\mathcal{P}\int_0^\infty \frac{\omega'\chi_i(\omega')}{\omega'^2-\omega^2}d\omega' \\ \chi_i(\omega) &= -\frac{2}{\pi}\mathcal{P}\int_0^\infty \frac{\omega\chi_r(\omega')}{\omega'^2-\omega^2}d\omega'\end{aligned} \tag{1.10.9}$$

There are several other ways to prove the Kramers-Kronig relations. For example, a more direct way is to state the causality condition in terms of the signum function $\text{sign}(t)$. Indeed, because $u(t)= \big(1+\text{sign}(t)\big)/2$, Eq. (1.10.4) may be written in the equivalent form $\chi(t)= \chi(t)\text{sign}(t)$. Then, Eq. (1.10.7) follows by applying the same frequency-domain convolution argument using the Fourier transform pair:

$$\text{sign}(t) \quad\Leftrightarrow\quad \mathcal{P}\frac{2}{j\omega} \tag{1.10.10}$$

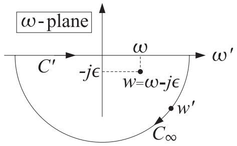
Alternatively, the causality condition can be expressed as $u(-t)\chi(t)= 0$. This approach is explored in Problem 1.12. Another proof is based on the *analyticity* properties of $\chi(\omega)$. Because of the causality condition, the Fourier integral in (1.10.3) can be restricted to the time range $0 < t < \infty$:

$$\chi(\omega)= \int_{-\infty}^{\infty} e^{-j\omega t}\chi(t)\,dt = \int_0^{\infty} e^{-j\omega t}\chi(t)\,dt \tag{1.10.11}$$

This implies that $\chi(\omega)$ can be analytically continued into the *lower* half ω-plane, so that replacing ω by $w = \omega - j\alpha$ with $\alpha \geq 0$ still gives a convergent Fourier integral in Eq. (1.10.11). Any singularities in $\chi(\omega)$ lie in the upper-half plane. For example, the simple model of Eq. (1.9.9) has poles at $\omega = \pm\bar{\omega}_0 + j\gamma/2$, where $\bar{\omega}_0 = \sqrt{\omega_0^2 - \gamma^2/4}$.

Next, we consider a clockwise closed contour $C = C' + C_\infty$ consisting of the real axis C' and an infinite semicircle C_∞ in the lower half-plane. Because $\chi(\omega)$ is analytic in the region enclosed by C, Cauchy's integral theorem implies that for any point w enclosed by C, that is, lying in the lower half-plane, we must have:

$$\chi(w)= -\frac{1}{2\pi j}\oint_C \frac{\chi(w')}{w'-w}\,dw' \tag{1.10.12}$$



where the overall minus sign arises because C was taken to be clockwise. Assuming that $\chi(\omega)$ falls off sufficiently fast for large ω, the contribution of the infinite semicircle can be ignored, thus leaving only the integral over the real axis. Setting $w = \omega - j\epsilon$ and taking the limit $\epsilon \to 0+$, we obtain the identical relationship to Eq. (1.10.5):

$$\chi(\omega)= -\lim_{\epsilon\to 0+}\frac{1}{2\pi j}\int_{-\infty}^{\infty}\frac{\chi(\omega')}{\omega'-\omega+j\epsilon}\,d\omega' = \frac{1}{2\pi}\int_{-\infty}^{\infty}\chi(\omega')\lim_{\epsilon\to 0+}\frac{1}{j(\omega-\omega')+\epsilon}\,d\omega'$$

An interesting consequence of the Kramers-Kronig relations is that there cannot exist a dielectric medium that is purely lossless, that is, such that $\chi_i(\omega)= 0$ for all ω, because this would also require that $\chi_r(\omega)= 0$ for all ω.

However, in all materials, $\chi_i(\omega)$ is significantly non-zero only in the neighborhoods of the medium's resonant frequencies, as for example in Fig. 1.9.1. In the frequency bands that are sufficiently far from the resonant bands, $\chi_i(\omega)$ may be assumed to be essentially zero. Such frequency bands are called *transparency bands* [130].

1.11 Group Velocity, Energy Velocity

Assuming a nonmagnetic material ($\mu = \mu_0$), a complex-valued refractive index may be defined by:

$$n(\omega)= n_r(\omega)-jn_i(\omega)= \sqrt{1+\chi(\omega)} = \sqrt{\frac{\epsilon(\omega)}{\epsilon_0}} \tag{1.11.1}$$

where n_r, n_i are its real and imaginary parts. Setting $\chi = \chi_r - j\chi_i$ we have the condition $n_r - jn_i = \sqrt{1+\chi_r - j\chi_i}$. Upon squaring, this splits into the two real-valued equations $n_r^2 - n_i^2 = 1 + \chi_r$ and $2n_r n_i = \chi_i$, with solutions:

$$\begin{aligned} n_r &= \left[\frac{\sqrt{(1+\chi_r)^2+\chi_i^2} + (1+\chi_r)}{2}\right]^{1/2} \\ n_i &= \text{sign}(\chi_i)\left[\frac{\sqrt{(1+\chi_r)^2+\chi_i^2} - (1+\chi_r)}{2}\right]^{1/2} = \frac{\chi_i}{2n_r}\end{aligned} \tag{1.11.2}$$

This form preserves the sign of χ_i, that is, n_i and χ_i are both positive for absorbing media, or both negative for gain media. The following approximate solution is often used, which can be justified whenever $|\chi| \ll 1$ (for example, in gases):

$$n_r(\omega)-jn_r(\omega)= \sqrt{1+\chi(\omega)} \simeq 1 + \frac{\chi}{2} \quad\Rightarrow\quad n_r = 1 + \frac{1}{2}\chi_r\,, \quad n_i = \frac{1}{2}\chi_i \tag{1.11.3}$$

We will see in Chap. 2 that a single-frequency uniform plane wave propagating, say, in the positive z-direction, has a wavenumber $k = \omega n/c = \omega(n_r - jn_i)/c \equiv \beta - j\alpha$, where c is the speed of light in vacuum. Therefore, the wave will have a space-time dependence:

$$e^{j(\omega t - kz)} = e^{j(\omega t-(\beta - j\alpha)z)} = e^{-\alpha z}e^{j(\omega t - \beta z)} = e^{-\omega n_i z/c}e^{j\omega(t - n_r z/c)} \tag{1.11.4}$$

The real part n_r defines the *phase velocity* of the wave, $v_p = \omega/\beta = c/n_r$, whereas the imaginary part n_i, or $\alpha = \omega n_i/c$, corresponds to attenuation or gain depending on the sign of n_i or χ_i.

When several such plane waves are superimposed to form a propagating pulse, we will see in Sec. 3.5 that the peak of the pulse (i.e., the point on the pulse where all the individual frequency components add up in phase), propagates with the so-called *group velocity* defined by:

$$v_g = \frac{d\omega}{d\beta} = \frac{1}{\dfrac{d\beta}{d\omega}} = \frac{c}{\dfrac{d(\omega n_r)}{d\omega}} = \frac{c}{n_r + \omega\dfrac{dn_r}{d\omega}} = \text{group velocity} \tag{1.11.5}$$

A *group refractive index* may be defined through $v_g = c/n_g$, or, $n_g = c/v_g$:

$$n_g = \frac{d(\omega n_r)}{d\omega} = n_r + \omega\frac{dn_r}{d\omega} = n_r - \lambda\frac{dn_r}{d\lambda} = \text{group refractive index} \qquad (1.11.6)$$

where λ is the free-space wavelength related to ω by $\lambda = 2\pi c/\omega$, and we used the differentiation property that $\omega d/d\omega = -\lambda d/d\lambda$.

Within an *anomalous* dispersion region, n_r is decreasing rapidly with ω, that is, $dn_r/d\omega < 0$, as in Fig. 1.9.1. This results in a group velocity v_g, given by Eq. (1.11.5), that may be larger than c or even negative. Such velocities are called "superluminal." Light pulses propagating at superluminal group velocities are referred to as "fast light" and we discuss them further in Sec. 3.9.

Within a *normal* dispersion region (e.g., to the far left and far right of the resonant frequency ω_0 in Fig. 1.9.1), n_r is an increasing function of ω, $dn_r/d\omega > 0$, which results in $v_g < c$. In specially engineered materials such as those exhibiting "electromagnetically induced transparency," the slope $dn_r/d\omega$ may be made so steep that the resulting group velocity v_g becomes extremely small, $v_g \ll c$. This is referred to as "slow light."

We close this section by showing that for lossless dispersive media, the energy velocity of a plane wave is equal to the group velocity defined by (1.11.5). This result is quite general, regardless of the frequency dependence of $\epsilon(\omega)$ and $\mu(\omega)$ (as long as these quantities are real.)

We will see in the next chapter that a plane wave propagating along the z-direction has electric and magnetic fields that are transverse to the z-direction and are related by:

$$|\boldsymbol{H}| = \frac{1}{\eta}|\boldsymbol{E}|\,, \quad \eta = \sqrt{\frac{\mu}{\epsilon}}$$

Moreover the time-averaged energy flux (in the z-direction) and energy density are:

$$\bar{\mathcal{P}}_z = \frac{|\boldsymbol{E}|^2}{2\eta}\,, \quad \bar{w} = \frac{1}{4}\frac{d(\omega\epsilon)}{d\omega}|\boldsymbol{E}|^2 + \frac{1}{4}\frac{d(\omega\mu)}{d\omega}|\boldsymbol{H}|^2 = \frac{1}{4}\left[\frac{d(\omega\epsilon)}{d\omega} + \frac{1}{\eta^2}\frac{d(\omega\mu)}{d\omega}\right]|\boldsymbol{E}|^2$$

The energy velocity is defined by $v_{\text{en}} = \bar{\mathcal{P}}_z/\bar{w}$. Thus, we have:

$$v_{\text{en}}^{-1} = \frac{\bar{w}}{\bar{\mathcal{P}}_z} = \frac{1}{2}\left[\eta\frac{d(\omega\epsilon)}{d\omega} + \frac{1}{\eta}\frac{d(\omega\mu)}{d\omega}\right] = \frac{1}{2}\left[\sqrt{\frac{\mu}{\epsilon}}\frac{d(\omega\epsilon)}{d\omega} + \sqrt{\frac{\epsilon}{\mu}}\frac{d(\omega\mu)}{d\omega}\right]$$

It is easily verified that the right-hand side can be expressed in terms of the wavenumber $k = \omega\sqrt{\epsilon\mu}$ in the form:

$$v_{\text{en}}^{-1} = \frac{1}{2}\left[\sqrt{\frac{\mu}{\epsilon}}\frac{d(\omega\epsilon)}{d\omega} + \sqrt{\frac{\epsilon}{\mu}}\frac{d(\omega\mu)}{d\omega}\right] = \frac{d(\omega\sqrt{\epsilon\mu})}{d\omega} = \frac{dk}{d\omega} = v_g^{-1} \qquad (1.11.7)$$

which shows the equality of the energy and group velocities. See Refs. [130–144] for further discussion on this topic.

Eq. (1.11.7) is also valid for the case of lossless negative-index media and implies that the group velocity, and hence the group refractive index $n_g = c_0/v_g$, will be positive, even though the refractive index n is negative. Writing $\epsilon = -|\epsilon|$ and $\mu = -|\mu|$ in this case and noting that $\eta = \sqrt{|\mu|/|\epsilon|}$ and $n = -\sqrt{|\epsilon\mu|}/\sqrt{\epsilon_0\mu_0}$, and $k = \omega n/c_0$, we have:

$$\begin{aligned} v_{\text{en}}^{-1} &= \frac{1}{2}\left[\sqrt{\frac{|\mu|}{|\epsilon|}}\frac{d(\omega\epsilon)}{d\omega} + \sqrt{\frac{|\epsilon|}{|\mu|}}\frac{d(\omega\mu)}{d\omega}\right] = -\frac{1}{2}\left[\sqrt{\frac{|\mu|}{|\epsilon|}}\frac{d(\omega|\epsilon|)}{d\omega} + \sqrt{\frac{|\epsilon|}{|\mu|}}\frac{d(\omega|\mu|)}{d\omega}\right] \\ &= -\frac{d(\omega\sqrt{|\epsilon\mu|})}{d\omega} = \frac{1}{c_0}\frac{d(\omega n)}{d\omega} = \frac{dk}{d\omega} = v_g^{-1} \end{aligned}$$

from which we also obtain the usual relationship $n_g = d(\omega n)/d\omega$. The positivity of v_g and n_g follows from the positivity of the derivatives $d(\omega\epsilon)/d\omega$ and $d(\omega\mu)/d\omega$, as required to keep v_{en} positive in negative-index media [353].

1.12 Problems

1.1 Prove the vector algebra identities:

$$\begin{array}{ll} \boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A}\cdot\boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A}\cdot\boldsymbol{B}) & \text{(BAC-CAB identity)} \\ \boldsymbol{A}\cdot(\boldsymbol{B}\times\boldsymbol{C}) = \boldsymbol{B}\cdot(\boldsymbol{C}\times\boldsymbol{A}) = \boldsymbol{C}\cdot(\boldsymbol{A}\times\boldsymbol{B}) & \\ |\boldsymbol{A}\times\boldsymbol{B}|^2 + |\boldsymbol{A}\cdot\boldsymbol{B}|^2 = |\boldsymbol{A}|^2|\boldsymbol{B}|^2 & \\ \boldsymbol{A} = \hat{\mathbf{n}}\times\boldsymbol{A}\times\hat{\mathbf{n}} + (\hat{\mathbf{n}}\cdot\boldsymbol{A})\hat{\mathbf{n}} & (\hat{\mathbf{n}}\text{ is any unit vector}) \end{array}$$

In the last identity, does it a make a difference whether $\hat{\mathbf{n}}\times\boldsymbol{A}\times\hat{\mathbf{n}}$ is taken to mean $\hat{\mathbf{n}}\times(\boldsymbol{A}\times\hat{\mathbf{n}})$ or $(\hat{\mathbf{n}}\times\boldsymbol{A})\times\hat{\mathbf{n}}$?

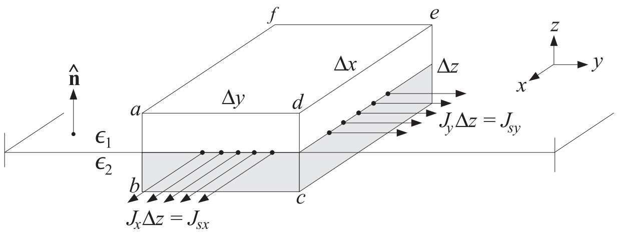
1.2 Prove the vector analysis identities:

$$\begin{array}{ll} \nabla\times(\nabla\phi) = 0 & \\ \nabla\cdot(\phi\nabla\psi) = \phi\nabla^2\psi + \nabla\phi\cdot\nabla\psi & \text{(Green's first identity)} \\ \nabla\cdot(\phi\nabla\psi - \psi\nabla\phi) = \phi\nabla^2\psi - \psi\nabla^2\phi & \text{(Green's second identity)} \\ \nabla\cdot(\phi\boldsymbol{A}) = (\nabla\phi)\cdot\boldsymbol{A} + \phi\nabla\cdot\boldsymbol{A} & \\ \nabla\times(\phi\boldsymbol{A}) = (\nabla\phi)\times\boldsymbol{A} + \phi\nabla\times\boldsymbol{A} & \\ \nabla\cdot(\nabla\times\boldsymbol{A}) = 0 & \\ \nabla\cdot\boldsymbol{A}\times\boldsymbol{B} = \boldsymbol{B}\cdot(\nabla\times\boldsymbol{A}) - \boldsymbol{A}\cdot(\nabla\times\boldsymbol{B}) & \\ \nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} & \end{array}$$

1.3 Consider the infinitesimal volume element $\Delta x\Delta y\Delta z$ shown below, such that its upper half lies in medium ϵ_1 and its lower half in medium ϵ_2. The axes are oriented such that $\hat{\mathbf{n}} = \hat{\mathbf{z}}$. Applying the integrated form of Ampère's law to the infinitesimal face $abcd$, show that

$$H_{2y} - H_{1y} = J_x\Delta z + \frac{\partial D_x}{\partial t}\Delta z$$

In the limit $\Delta z \to 0$, the second term in the right-hand side may be assumed to go to zero, whereas the first term will be non-zero and may be set equal to a surface current density, that is, $J_{sx} \equiv \lim_{\Delta z\to 0}(J_x\Delta z)$. Show that this leads to the boundary condition $H_{1y} - H_{2y} = -J_{sx}$. Similarly, show that $H_{1x} - H_{2x} = J_{sy}$, and that these two boundary conditions can be combined vectorially into Eq. (1.4.4).



Next, apply the integrated form of Gauss's law to the same volume element and show the boundary condition: $D_{1z} - D_{2z} = \rho_s = \lim_{\Delta z\to 0}(\rho\Delta z)$.

1.4 Show that the time average of the *product* of two harmonic quantities $\mathcal{A}(t) = \mathrm{Re}[Ae^{j\omega t}]$ and $\mathcal{B}(t) = \mathrm{Re}[Be^{j\omega t}]$ with phasors A, B is given by:

$$\overline{\mathcal{A}(t)\mathcal{B}(t)} = \frac{1}{T}\int_0^T \mathcal{A}(t)\mathcal{B}(t)\,dt = \frac{1}{2}\mathrm{Re}[AB^*]$$

where $T = 2\pi/\omega$ is one period. Then, show that the time-averaged values of the cross and dot products of two time-harmonic vector quantities $\boldsymbol{\mathcal{A}}(t) = \mathrm{Re}[\boldsymbol{A}e^{j\omega t}]$ and $\boldsymbol{\mathcal{B}}(t) = \mathrm{Re}[\boldsymbol{B}e^{j\omega t}]$ can be expressed in terms of the corresponding phasors as follows:

$$\overline{\boldsymbol{\mathcal{A}}(t)\times\boldsymbol{\mathcal{B}}(t)} = \frac{1}{2}\mathrm{Re}[\boldsymbol{A}\times\boldsymbol{B}^*], \qquad \overline{\boldsymbol{\mathcal{A}}(t)\cdot\boldsymbol{\mathcal{B}}(t)} = \frac{1}{2}\mathrm{Re}[\boldsymbol{A}\cdot\boldsymbol{B}^*]$$

1.5 Assuming that $\boldsymbol{B} = \mu\boldsymbol{H}$, show that Maxwell's equations (1.8.2) imply the following complex-valued version of Poynting's theorem:

$$\nabla\cdot(\boldsymbol{E}\times\boldsymbol{H}^*) = -j\omega\mu\boldsymbol{H}\cdot\boldsymbol{H}^* - \boldsymbol{E}\cdot\boldsymbol{J}_{\mathrm{tot}}^*, \qquad \text{where } \boldsymbol{J}_{\mathrm{tot}} = \boldsymbol{J} + j\omega\boldsymbol{D}$$

Extracting the real-parts of both sides and integrating over a volume V bounded by a closed surface S, show the time-averaged form of energy conservation:

$$-\oint_S \frac{1}{2}\mathrm{Re}[\boldsymbol{E}\times\boldsymbol{H}^*]\cdot d\boldsymbol{S} = \int_V \frac{1}{2}\mathrm{Re}[\boldsymbol{E}\cdot\boldsymbol{J}_{\mathrm{tot}}^*]\,dV$$

which states that the net time-averaged power flowing into a volume is dissipated into heat. For a lossless dielectric, show that the above integrals are zero and provide an interpretation.

1.6 Assuming that $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$, show that Maxwell's equations (1.1.1) imply the following relationships:

$$\rho E_x + \Big(\boldsymbol{D}\times\frac{\partial\boldsymbol{B}}{\partial t}\Big)_x = \nabla\cdot\Big(\epsilon E_x\boldsymbol{E} - \hat{\mathbf{x}}\frac{1}{2}\epsilon E^2\Big)$$

$$(\boldsymbol{J}\times\boldsymbol{B})_x + \Big(\frac{\partial\boldsymbol{D}}{\partial t}\times\boldsymbol{B}\Big)_x = \nabla\cdot\Big(\mu H_x\boldsymbol{H} - \hat{\mathbf{x}}\frac{1}{2}\mu H^2\Big)$$

where the subscript x means the x-component. From these, derive the following relationship that represents momentum conservation:

$$f_x + \frac{\partial G_x}{\partial t} = \nabla\cdot\boldsymbol{T}_x \tag{1.12.1}$$

where f_x, G_x are the x-components of the vectors $\boldsymbol{f} = \rho\boldsymbol{E} + \boldsymbol{J}\times\boldsymbol{B}$ and $\boldsymbol{G} = \boldsymbol{D}\times\boldsymbol{B}$, and $\boldsymbol{T}_x$ is defined to be the vector (equal to Maxwell's stress tensor acting on the unit vector $\hat{\mathbf{x}}$):

$$\boldsymbol{T}_x = \epsilon E_x\boldsymbol{E} + \mu H_x\boldsymbol{H} - \hat{\mathbf{x}}\frac{1}{2}(\epsilon E^2 + \mu H^2)$$

Write similar equations of the y, z components. The quantity G_x is interpreted as the field momentum (in the x-direction) per unit volume, that is, the momentum density.

1.7 Show that the causal and stable time-domain dielectric response corresponding to Eq. (1.9.7) is given as follows:

$$\epsilon(t) = \epsilon_0\delta(t) + \epsilon_0\chi(t), \quad \chi(t) = \frac{\omega_p^2}{\bar{\omega}_0}e^{-\gamma t/2}\sin(\bar{\omega}_0 t)\,u(t) \tag{1.12.2}$$

where $u(t)$ is the unit-step function and $\bar{\omega}_0 = \sqrt{\omega_0^2 - \gamma^2/4}$, and we must assume that $\gamma < 2\omega_0$, as is typically the case in practice. Discuss the solution for the case $\gamma/2 > \omega_0$.

1.8 Show that the plasma frequency for electrons can be expressed in the simple numerical form: $f_p = 9\sqrt{N}$, where f_p is in Hz and N is the electron density in electrons/m^3. What is f_p for the ionosphere if $N = 10^{12}$? [*Ans.* 9 MHz.]

1.9 Show that the relaxation equation (1.9.26) can be written in the following form in terms of the dc-conductivity σ defined by Eq. (1.9.22):

$$\frac{1}{\gamma}\ddot{\rho}(\mathbf{r},t) + \dot{\rho}(\mathbf{r},t) + \frac{\sigma}{\epsilon_0}\rho(\mathbf{r},t) = 0$$

Then, show that it reduces to the naive relaxation equation (1.6.3) in the limit $\tau = 1/\gamma \to 0$. Show also that in this limit, Ohm's law (1.9.25) takes the instantaneous form $\boldsymbol{J} = \sigma\boldsymbol{E}$, from which the naive relaxation constant $\tau_{\mathrm{relax}} = \epsilon_0/\sigma$ was derived.

1.10 Conductors and plasmas exhibit anisotropic and birefringent behavior when they are in the presence of an external magnetic field. The equation of motion of conduction electrons in a constant external magnetic field is $m\dot{\boldsymbol{v}} = e(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}) - m\gamma\boldsymbol{v}$, with the collisional term included. Assume the magnetic field is in the z-direction, $\boldsymbol{B} = \hat{\mathbf{z}}B$, and that $\boldsymbol{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y$ and $\boldsymbol{v} = \hat{\mathbf{x}}v_x + \hat{\mathbf{y}}v_y$.

a. Show that in component form, the above equations of motion read:

$$\begin{aligned}\dot{v}_x &= \frac{e}{m}E_x + \omega_B v_y - \gamma v_x \\ \dot{v}_y &= \frac{e}{m}E_y - \omega_B v_x - \gamma v_y\end{aligned} \qquad \text{where} \quad \omega_B = \frac{eB}{m} = \text{(cyclotron frequency)}$$

What is the cyclotron frequency in Hz for electrons in the Earth's magnetic field $B = 0.4$ gauss $= 0.4\times10^{-4}$ Tesla? [*Ans.* 1.12 MHz.]

b. To solve this system, work with the combinations $v_x \pm jv_y$. Assuming harmonic time-dependence, show that the solution is:

$$v_x \pm jv_y = \frac{\dfrac{e}{m}(E_x \pm jE_y)}{\gamma + j(\omega \pm \omega_B)}$$

c. Define the induced currents as $\boldsymbol{J} = Ne\boldsymbol{v}$. Show that:

$$J_x \pm jJ_y = \sigma_\pm(\omega)\,(E_x \pm jE_y), \qquad \text{where} \quad \sigma_\pm(\omega) = \frac{\gamma\sigma_0}{\gamma + j(\omega \pm \omega_B)}$$

where $\sigma_0 = \dfrac{Ne^2}{m\gamma}$ is the dc value of the conductivity.

d. Show that the t-domain version of part (c) is:

$$J_x(t)\pm jJ_y(t)=\int_0^t \sigma_\pm(t-t')\big(E_x(t')\pm jE_y(t')\big)dt'$$

where $\sigma_\pm(t)=\gamma\sigma_0 e^{-\gamma t}e^{\mp j\omega_B t}u(t)$ is the inverse Fourier transform of $\sigma_\pm(\omega)$ and $u(t)$ is the unit-step function.

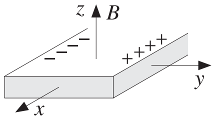
e. Rewrite part (d) in component form:

$$\begin{aligned}J_x(t) &= \int_0^t \big[\sigma_{xx}(t-t')E_x(t')+\sigma_{xy}(t-t')E_y(t')\big]dt'\\ J_y(t) &= \int_0^t \big[\sigma_{yx}(t-t')E_x(t')+\sigma_{yy}(t-t')E_y(t')\big]dt'\end{aligned}$$

and identify the quantities $\sigma_{xx}(t),\sigma_{xy}(t),\sigma_{yx}(t),\sigma_{yy}(t)$.

f. Evaluate part (e) in the special case $E_x(t)=E_x u(t)$ and $E_y(t)=E_y u(t)$, where E_x, E_y are constants, and show that after a long time the steady-state version of part (e) will be:

$$\begin{aligned}J_x &= \sigma_0\frac{E_x+bE_y}{1+b^2}\\ J_y &= \sigma_0\frac{E_y-bE_x}{1+b^2}\end{aligned}$$



where $b=\omega_B/\gamma$. If the conductor has finite extent in the y-direction, as shown above, then no steady current can flow in this direction, $J_y=0$. This implies that if an electric field is applied in the x-direction, an electric field will develop across the y-ends of the conductor, $E_y=bE_x$. The conduction charges will tend to accumulate either on the right or the left side of the conductor, depending on the sign of b, which depends on the sign of the electric charge e. This is the *Hall effect* and is used to determine the sign of the conduction charges in semiconductors, e.g., positive holes for p-type, or negative electrons for n-type.

What is the numerical value of b for electrons in copper if B is 1 gauss? [*Ans.* 43.]

g. For a collisionless plasma ($\gamma=0$), show that its dielectric behavior is determined from $D_x\pm jD_y=\epsilon_\pm(\omega)(E_x\pm jE_y)$, where

$$\epsilon_\pm(\omega)=\epsilon_0\left(1-\frac{\omega_p^2}{\omega(\omega\pm\omega_B)}\right)$$

where ω_p is the plasma frequency. Thus, the plasma exhibits birefringence.

1.11 This problem deals with various properties of the Kramers-Kronig dispersion relations for the electric susceptibility, given by Eq. (1.10.8).

a. Using the symmetry properties $\chi_r(\omega)=\chi_r(-\omega)$ and $\chi_i(\omega)=-\chi_i(-\omega)$, show that (1.10.8) can be written in the folded form of Eq. (1.10.9).

b. Using the definition of principal-value integrals, show the following integral:

$$\mathcal{P}\int_0^\infty \frac{d\omega'}{\omega'^2-\omega^2}=0 \tag{1.12.3}$$

Hint: You may use the following indefinite integral: $\displaystyle\int\frac{dx}{a^2-x^2}=\frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right|$.

c. Using Eq. (1.12.3), show that the relations (1.10.9) may be rewritten as ordinary integrals (without the $\mathcal{P}$ instruction) as follows:

$$\begin{aligned}\chi_r(\omega) &= \frac{2}{\pi}\int_0^\infty\frac{\omega'\chi_i(\omega')-\omega\chi_i(\omega)}{\omega'^2-\omega^2}\,d\omega'\\ \chi_i(\omega) &= -\frac{2}{\pi}\int_0^\infty\frac{\omega\chi_r(\omega')-\omega\chi_r(\omega)}{\omega'^2-\omega^2}\,d\omega'\end{aligned} \tag{1.12.4}$$

Hint: You will need to argue that the integrands have no singularity at $\omega'=\omega$.

d. For a simple oscillator model of dielectric polarization, the susceptibility is given by:

$$\begin{aligned}\chi(\omega) &= \chi_r(\omega)-j\chi_i(\omega)=\frac{\omega_p^2}{\omega_0^2-\omega^2+j\gamma\omega}\\ &= \frac{\omega_p^2(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}-j\frac{\gamma\omega\omega_p^2}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}\end{aligned} \tag{1.12.5}$$

Show that for this model the quantities $\chi_r(\omega)$ and $\chi_i(\omega)$ satisfy the modified Kramers-Kronig relationships (1.12.4). *Hint*: You may use the following definite integrals, for which you may assume that $0<\gamma<2\omega_0$:

$$\frac{2}{\pi}\int_0^\infty\frac{dx}{(\omega_0^2-x^2)^2+\gamma^2x^2}=\frac{1}{\gamma\omega_0^2}\,,\qquad \frac{2}{\pi}\int_0^\infty\frac{x^2dx}{(\omega_0^2-x^2)^2+\gamma^2x^2}=\frac{1}{\gamma}$$

Indeed, show that these integrals may be reduced to the following ones, which can be found in standard tables of integrals:

$$\frac{2}{\pi}\int_0^\infty\frac{dy}{1-2y^2\cos\theta+y^4}=\frac{2}{\pi}\int_0^\infty\frac{y^2dy}{1-2y^2\cos\theta+y^4}=\frac{1}{\sqrt{2(1-\cos\theta)}}$$

where $\sin(\theta/2)=\gamma/(2\omega_0)$.

e. Consider the limit of Eq. (1.12.5) as $\gamma\to 0$. Show that in this case the functions χ_r,χ_i are given as follows, and that they still satisfy the Kramers-Kronig relations:

$$\chi_r(\omega)=\mathcal{P}\frac{\omega_p^2}{\omega_0-\omega}+\mathcal{P}\frac{\omega_p^2}{\omega_0+\omega}\,,\quad \chi_i(\omega)=\frac{\pi\omega_p^2}{2\omega_0}\big[\delta(\omega-\omega_0)-\delta(\omega+\omega_0)\big]$$

1.12 Derive the Kramers-Kronig relationship of Eq. (1.10.7) by starting with the causality condition $\chi(t)u(-t)=0$ and translating it to the frequency domain, that is, expressing it as the convolution of the Fourier transforms of $\chi(t)$ and $u(-t)$.

1.13 An isotropic homogeneous lossless dielectric medium is moving with uniform velocity $\boldsymbol{v}$ with respect to a fixed coordinate frame S. In the frame S' moving with dielectric, the constitutive relations are assumed to be the usual ones, that is, $\boldsymbol{D}'=\epsilon\boldsymbol{E}'$ and $\boldsymbol{B}'=\mu\boldsymbol{H}'$. Using the Lorentz transformations given in Eq. (H.30) of Appendix H, show that the constitutive relations take the following form in the fixed frame S:

$$\boldsymbol{D}=\epsilon\boldsymbol{E}+a\boldsymbol{v}\times(\boldsymbol{H}-\epsilon\boldsymbol{v}\times\boldsymbol{E})\,,\quad \boldsymbol{B}=\mu\boldsymbol{H}-a\boldsymbol{v}\times(\boldsymbol{E}+\mu\boldsymbol{v}\times\boldsymbol{H})\,,\quad a=\frac{\epsilon\mu-\epsilon_0\mu_0}{1-\epsilon\mu v^2}$$

2

Uniform Plane Waves

2.1 Uniform Plane Waves in Lossless Media

The simplest electromagnetic waves are uniform plane waves propagating along some fixed direction, say the z-direction, in a lossless medium $\{\epsilon, \mu\}$.

The assumption of uniformity means that the fields have no dependence on the transverse coordinates x, y and are functions only of z, t. Thus, we look for solutions of Maxwell's equations of the form: $\boldsymbol{E}(x,y,z,t) = \boldsymbol{E}(z,t)$ and $\boldsymbol{H}(x,y,z,t) = \boldsymbol{H}(z,t)$.

Because there is no dependence on x, y, we set the partial derivatives[†] $\partial_x = 0$ and $\partial_y = 0$. Then, the gradient, divergence, and curl operations take the simplified forms:

$$\nabla = \hat{\mathbf{z}}\,\frac{\partial}{\partial z}\,, \quad \nabla\cdot\boldsymbol{E} = \frac{\partial E_z}{\partial z}\,, \quad \nabla\times\boldsymbol{E} = \hat{\mathbf{z}}\times\frac{\partial \boldsymbol{E}}{\partial z} = -\hat{\mathbf{x}}\,\frac{\partial E_y}{\partial z} + \hat{\mathbf{y}}\,\frac{\partial E_x}{\partial z}$$

Assuming that $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$, the source-free Maxwell's equations become:

$$\begin{aligned}
\nabla\times\boldsymbol{E} &= -\mu\,\frac{\partial \boldsymbol{H}}{\partial t} \\
\nabla\times\boldsymbol{H} &= \epsilon\,\frac{\partial \boldsymbol{E}}{\partial t} \\
\nabla\cdot\boldsymbol{E} &= 0 \\
\nabla\cdot\boldsymbol{H} &= 0
\end{aligned}
\quad\Rightarrow\quad
\begin{aligned}
\hat{\mathbf{z}}\times\frac{\partial \boldsymbol{E}}{\partial z} &= -\mu\,\frac{\partial \boldsymbol{H}}{\partial t} \\
\hat{\mathbf{z}}\times\frac{\partial \boldsymbol{H}}{\partial z} &= \epsilon\,\frac{\partial \boldsymbol{E}}{\partial t} \\
\frac{\partial E_z}{\partial z} &= 0 \\
\frac{\partial H_z}{\partial z} &= 0
\end{aligned}
\tag{2.1.1}$$

An immediate consequence of uniformity is that $\boldsymbol{E}$ and $\boldsymbol{H}$ do not have components along the z-direction, that is, $E_z = H_z = 0$. Taking the dot-product of Ampère's law with the unit vector $\hat{\mathbf{z}}$, and using the identity $\hat{\mathbf{z}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{A}) = 0$, we have:

$$\hat{\mathbf{z}}\cdot\left(\hat{\mathbf{z}}\times\frac{\partial \boldsymbol{H}}{\partial z}\right) = \epsilon\,\hat{\mathbf{z}}\cdot\frac{\partial \boldsymbol{E}}{\partial t} = 0 \quad\Rightarrow\quad \frac{\partial E_z}{\partial t} = 0$$

[†] The shorthand notation ∂_x stands for $\dfrac{\partial}{\partial x}$.

Because also $\partial_z E_z = 0$, it follows that E_z must be a constant, independent of z, t. Excluding static solutions, we may take this constant to be zero. Similarly, we have $H_z = 0$. Thus, the fields have components only along the x, y directions:

$$\begin{aligned} \mathbf{E}(z,t) &= \hat{\mathbf{x}}\,E_x(z,t) + \hat{\mathbf{y}}\,E_y(z,t) \\ \mathbf{H}(z,t) &= \hat{\mathbf{x}}\,H_x(z,t) + \hat{\mathbf{y}}\,H_y(z,t) \end{aligned} \qquad \text{(transverse fields)} \qquad (2.1.2)$$

These fields must satisfy Faraday's and Ampère's laws in Eqs. (2.1.1). We rewrite these equations in a more convenient form by replacing ϵ and μ by:

$$\epsilon = \frac{1}{\eta c}\,, \quad \mu = \frac{\eta}{c}\,, \qquad \text{where} \quad c = \frac{1}{\sqrt{\mu\epsilon}}\,, \quad \eta = \sqrt{\frac{\mu}{\epsilon}} \qquad (2.1.3)$$

Thus, c, η are the speed of light and characteristic impedance of the propagation medium. Then, the first two of Eqs. (2.1.1) may be written in the equivalent forms:

$$\begin{aligned} \hat{\mathbf{z}} \times \frac{\partial \mathbf{E}}{\partial z} &= -\frac{1}{c}\,\eta\,\frac{\partial \mathbf{H}}{\partial t} \\ \eta\,\hat{\mathbf{z}} \times \frac{\partial \mathbf{H}}{\partial z} &= \frac{1}{c}\,\frac{\partial \mathbf{E}}{\partial t} \end{aligned} \qquad (2.1.4)$$

The first may be solved for $\partial_z \mathbf{E}$ by crossing it with $\hat{\mathbf{z}}$. Using the BAC-CAB rule, and noting that $\mathbf{E}$ has no z-component, we have:

$$\left(\hat{\mathbf{z}} \times \frac{\partial \mathbf{E}}{\partial z}\right) \times \hat{\mathbf{z}} = \frac{\partial \mathbf{E}}{\partial z}(\hat{\mathbf{z}}\cdot\hat{\mathbf{z}}) - \hat{\mathbf{z}}\left(\hat{\mathbf{z}} \cdot \frac{\partial \mathbf{E}}{\partial z}\right) = \frac{\partial \mathbf{E}}{\partial z}$$

where we used $\hat{\mathbf{z}} \cdot \partial_z \mathbf{E} = \partial_z E_z = 0$ and $\hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1$. It follows that Eqs. (2.1.4) may be replaced by the equivalent system:

$$\boxed{\begin{aligned} \frac{\partial \mathbf{E}}{\partial z} &= -\frac{1}{c}\,\frac{\partial}{\partial t}(\eta \mathbf{H} \times \hat{\mathbf{z}}) \\ \frac{\partial}{\partial z}(\eta \mathbf{H} \times \hat{\mathbf{z}}) &= -\frac{1}{c}\,\frac{\partial \mathbf{E}}{\partial t} \end{aligned}} \qquad (2.1.5)$$

Now all the terms have the same dimension. Eqs. (2.1.5) imply that both $\mathbf{E}$ and $\mathbf{H}$ satisfy the *one-dimensional wave equation*. Indeed, differentiating the first equation with respect to z and using the second, we have:

$$\frac{\partial^2 \mathbf{E}}{\partial z^2} = -\frac{1}{c}\,\frac{\partial}{\partial t}\,\frac{\partial}{\partial z}(\eta \mathbf{H} \times \hat{\mathbf{z}}) = \frac{1}{c^2}\,\frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad \text{or,}$$

$$\boxed{\left(\frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\,\frac{\partial^2}{\partial t^2}\right)\mathbf{E}(z,t) = 0} \qquad \text{(wave equation)} \qquad (2.1.6)$$

and similarly for $\mathbf{H}$. Rather than solving the wave equation, we prefer to work directly with the coupled system (2.1.5). The system can be decoupled by introducing the so-called *forward* and *backward* electric fields defined as the linear combinations:

$$\begin{aligned} \mathbf{E}_+ &= \frac{1}{2}(\mathbf{E} + \eta \mathbf{H} \times \hat{\mathbf{z}}) \\ \mathbf{E}_- &= \frac{1}{2}(\mathbf{E} - \eta \mathbf{H} \times \hat{\mathbf{z}}) \end{aligned} \qquad \text{(forward and backward fields)} \qquad (2.1.7)$$

Component-wise, these are:

$$E_{x\pm} = \frac{1}{2}(E_x \pm \eta H_y)\,, \quad E_{y\pm} = \frac{1}{2}(E_y \mp \eta H_x) \qquad (2.1.8)$$

We show next that $\mathbf{E}_+(z,t)$ corresponds to a forward-moving wave, that is, moving towards the positive z-direction, and $\mathbf{E}_-(z,t)$, to a backward-moving wave. Eqs. (2.1.7) can be inverted to express $\mathbf{E}, \mathbf{H}$ in terms of $\mathbf{E}_+, \mathbf{E}_-$. Adding and subtracting them, and using the BAC-CAB rule and the orthogonality conditions $\hat{\mathbf{z}} \cdot \mathbf{E}_\pm = 0$, we obtain:

$$\begin{aligned} \mathbf{E}(z,t) &= \mathbf{E}_+(z,t) + \mathbf{E}_-(z,t) \\ \mathbf{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}} \times \big[\mathbf{E}_+(z,t) - \mathbf{E}_-(z,t)\big] \end{aligned} \qquad (2.1.9)$$

In terms of the forward and backward fields $\mathbf{E}_\pm$, the system of Eqs. (2.1.5) *decouples* into two separate equations:

$$\boxed{\begin{aligned} \frac{\partial \mathbf{E}_+}{\partial z} &= -\frac{1}{c}\,\frac{\partial \mathbf{E}_+}{\partial t} \\ \frac{\partial \mathbf{E}_-}{\partial z} &= +\frac{1}{c}\,\frac{\partial \mathbf{E}_-}{\partial t} \end{aligned}} \qquad (2.1.10)$$

Indeed, using Eqs. (2.1.5), we verify:

$$\frac{\partial}{\partial z}(\mathbf{E} \pm \eta \mathbf{H} \times \hat{\mathbf{z}}) = -\frac{1}{c}\,\frac{\partial}{\partial t}(\eta \mathbf{H} \times \hat{\mathbf{z}}) \mp \frac{1}{c}\,\frac{\partial \mathbf{E}}{\partial t} = \mp\frac{1}{c}\,\frac{\partial}{\partial t}(\mathbf{E} \pm \eta \mathbf{H} \times \hat{\mathbf{z}})$$

Eqs. (2.1.10) can be solved by noting that the forward field $\mathbf{E}_+(z,t)$ must depend on z, t only through the combination $z - ct$ (for a proof, see Problem 2.1.) If we set $\mathbf{E}_+(z,t) = \mathbf{F}(z - ct)$, where $\mathbf{F}(\zeta)$ is an arbitrary function of its argument $\zeta = z - ct$, then we will have:

$$\begin{aligned} \frac{\partial \mathbf{E}_+}{\partial z} &= \frac{\partial}{\partial z}\mathbf{F}(z - ct) = \frac{\partial \zeta}{\partial z}\,\frac{\partial \mathbf{F}(\zeta)}{\partial \zeta} = \frac{\partial \mathbf{F}(\zeta)}{\partial \zeta} \\ \frac{\partial \mathbf{E}_+}{\partial t} &= \frac{\partial}{\partial t}\mathbf{F}(z - ct) = \frac{\partial \zeta}{\partial t}\,\frac{\partial \mathbf{F}(\zeta)}{\partial \zeta} = -c\,\frac{\partial \mathbf{F}(\zeta)}{\partial \zeta} \end{aligned} \quad \Rightarrow \quad \frac{\partial \mathbf{E}_+}{\partial z} = -\frac{1}{c}\,\frac{\partial \mathbf{E}_+}{\partial t}$$

Vectorially, $\mathbf{F}$ must have only x, y components, $\mathbf{F} = \hat{\mathbf{x}}F_x + \hat{\mathbf{y}}F_y$, that is, it must be transverse to the propagation direction, $\hat{\mathbf{z}} \cdot \mathbf{F} = 0$.

Similarly, we find from the second of Eqs. (2.1.10) that $\mathbf{E}_-(z,t)$ must depend on z, t through the combination $z + ct$, so that $\mathbf{E}_-(z,t) = \mathbf{G}(z + ct)$, where $\mathbf{G}(\xi)$ is an arbitrary (transverse) function of $\xi = z + ct$. In conclusion, the most general solutions for the forward and backward fields of Eqs. (2.1.10) are:

$$\begin{aligned} \mathbf{E}_+(z,t) &= \mathbf{F}(z - ct) \\ \mathbf{E}_-(z,t) &= \mathbf{G}(z + ct) \end{aligned} \qquad (2.1.11)$$

with arbitrary functions $\mathbf{F}$ and $\mathbf{G}$, such that $\hat{\mathbf{z}} \cdot \mathbf{F} = \hat{\mathbf{z}} \cdot \mathbf{G} = 0$.

Inserting these into the inverse formula (2.1.9), we obtain the most general solution of (2.1.5), expressed as a linear combination of forward and backward waves:

$$\boxed{\begin{aligned} \mathbf{E}(z,t) &= \mathbf{F}(z-ct)+\mathbf{G}(z+ct) \\ \mathbf{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\big[\mathbf{F}(z-ct)-\mathbf{G}(z+ct)\big] \end{aligned}} \tag{2.1.12}$$

The term $\mathbf{E}_+(z,t)=\mathbf{F}(z-ct)$ represents a wave propagating with speed c in the positive z-direction, while $\mathbf{E}_-(z,t)=\mathbf{G}(z+ct)$ represents a wave traveling in the negative z-direction.

To see this, consider the forward field at a later time $t+\Delta t$. During the time interval Δt, the wave moves in the positive z-direction by a distance $\Delta z = c\Delta t$. Indeed, we have:

$$\left.\begin{aligned} \mathbf{E}_+(z,t+\Delta t) &= \mathbf{F}\big(z-c(t+\Delta t)\big) = \mathbf{F}(z-c\Delta t-ct) \\ \mathbf{E}_+(z-\Delta z,t) &= \mathbf{F}\big((z-\Delta z)-ct\big) = \mathbf{F}(z-c\Delta t-ct) \end{aligned}\right\} \Rightarrow\ \mathbf{E}_+(z,t+\Delta t)=\mathbf{E}_+(z-\Delta z,t)$$

This states that the forward field at time $t+\Delta t$ is the same as the field at time t, but translated to the *right* along the z-axis by a distance $\Delta z = c\Delta t$. Equivalently, the field at location $z+\Delta z$ at time t is the same as the field at location z at the earlier time $t-\Delta t = t-\Delta z/c$, that is,

$$\mathbf{E}_+(z+\Delta z,t)=\mathbf{E}_+(z,t-\Delta t)$$

Similarly, we find that $\mathbf{E}_-(z,t+\Delta t)=\mathbf{E}_-(z+\Delta z,t)$, which states that the backward field at time $t+\Delta t$ is the same as the field at time t, translated to the *left* by a distance Δz. Fig. 2.1.1 depicts these two cases.

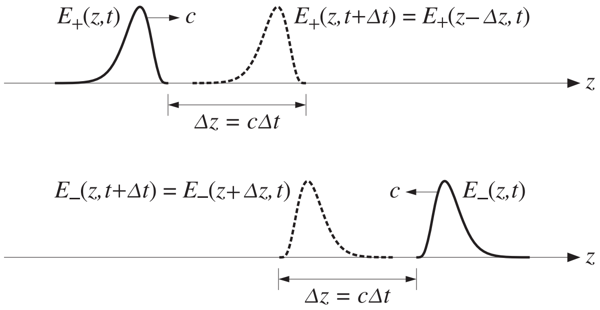
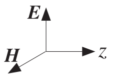

$E_+(z,t)$
c
$E_+(z,t+\Delta t) = E_+(z-\Delta z,t)$
z
$\Delta z = c\Delta t$
$E_-(z,t+\Delta t) = E_-(z+\Delta z,t)$
c
$E_-(z,t)$
z
$\Delta z = c\Delta t$

Fig. 2.1.1 Forward and backward waves.

The two special cases corresponding to forward waves only $(\mathbf{G}=0)$, or to backward ones $(\mathbf{F}=0)$, are of particular interest. For the forward case, we have:

$$\begin{aligned} \mathbf{E}(z,t) &= \mathbf{F}(z-ct) \\ \mathbf{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\mathbf{F}(z-ct)=\frac{1}{\eta}\,\hat{\mathbf{z}}\times\mathbf{E}(z,t) \end{aligned} \tag{2.1.13}$$



This solution has the following properties: (a) The field vectors $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to each other, $\mathbf{E}\cdot\mathbf{H}=0$, while they are transverse to the z-direction, (b) The three vectors $\{\mathbf{E},\mathbf{H},\hat{\mathbf{z}}\}$ form a *right-handed* vector system as shown in the figure, in the sense that $\mathbf{E}\times\mathbf{H}$ points in the direction of $\hat{\mathbf{z}}$, (c) The ratio of $\mathbf{E}$ to $\mathbf{H}\times\hat{\mathbf{z}}$ is independent of z,t and equals the characteristic impedance η of the propagation medium; indeed:

$$\mathbf{H}(z,t)=\frac{1}{\eta}\,\hat{\mathbf{z}}\times\mathbf{E}(z,t) \quad\Rightarrow\quad \mathbf{E}(z,t)=\eta\mathbf{H}(z,t)\times\hat{\mathbf{z}} \tag{2.1.14}$$

The electromagnetic energy of such forward wave flows in the positive z-direction. With the help of the BAC-CAB rule, we find for the Poynting vector:

$$\boldsymbol{\mathcal{P}} = \mathbf{E}\times\mathbf{H} = \hat{\mathbf{z}}\,\frac{1}{\eta}|\mathbf{F}|^2 = c\,\hat{\mathbf{z}}\,\epsilon|\mathbf{F}|^2 \tag{2.1.15}$$

where we denoted $|\mathbf{F}|^2=\mathbf{F}\cdot\mathbf{F}$ and replaced $1/\eta = c\epsilon$. The electric and magnetic energy densities (per unit volume) turn out to be equal to each other. Because $\hat{\mathbf{z}}$ and $\mathbf{F}$ are mutually orthogonal, we have for the cross product $|\hat{\mathbf{z}}\times\mathbf{F}| = |\hat{\mathbf{z}}||\mathbf{F}| = |\mathbf{F}|$. Then,

$$\begin{aligned} w_e &= \frac{1}{2}\epsilon\,|\mathbf{E}|^2 = \frac{1}{2}\epsilon|\mathbf{F}|^2 \\ w_m &= \frac{1}{2}\mu\,|\mathbf{H}|^2 = \frac{1}{2}\mu\,\frac{1}{\eta^2}|\hat{\mathbf{z}}\times\mathbf{F}|^2 = \frac{1}{2}\epsilon\,|\mathbf{F}|^2 = w_e \end{aligned}$$

where we replaced $\mu/\eta^2=\epsilon$. Thus, the total energy density of the forward wave will be:

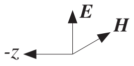
$$w = w_e + w_m = 2w_e = \epsilon|\mathbf{F}|^2 \tag{2.1.16}$$

In accordance with the flux/density relationship of Eq. (1.5.2), the transport velocity of the electromagnetic energy is found to be:

$$\mathbf{v} = \frac{\boldsymbol{\mathcal{P}}}{w} = \frac{c\,\hat{\mathbf{z}}\,\epsilon|\mathbf{F}|^2}{\epsilon|\mathbf{F}|^2} = c\,\hat{\mathbf{z}}$$

As expected, the energy of the forward-moving wave is being transported at a speed c along the positive z-direction. Similar results can be derived for the backward-moving solution that has $\mathbf{F}=0$ and $\mathbf{G}\neq 0$. The fields are now:

$$\begin{aligned} \mathbf{E}(z,t) &= \mathbf{G}(z+ct) \\ \mathbf{H}(z,t) &= -\frac{1}{\eta}\,\hat{\mathbf{z}}\times\mathbf{G}(z+ct)=-\frac{1}{\eta}\,\hat{\mathbf{z}}\times\mathbf{E}(z,t) \end{aligned} \tag{2.1.17}$$



The Poynting vector becomes $\boldsymbol{\mathcal{P}} = \mathbf{E}\times\mathbf{H} = -c\,\hat{\mathbf{z}}\,\epsilon|\mathbf{G}|^2$ and points in the negative z-direction, that is, the propagation direction. The energy transport velocity is $\mathbf{v}=-c\,\hat{\mathbf{z}}$. Now, the vectors $\{\mathbf{E},\mathbf{H},-\hat{\mathbf{z}}\}$ form a right-handed system, as shown. The ratio of E to H is still equal to η, provided we replace $\hat{\mathbf{z}}$ with $-\hat{\mathbf{z}}$:

$$\mathbf{H}(z,t)=\frac{1}{\eta}(-\hat{\mathbf{z}})\times\mathbf{E}(z,t) \quad\Rightarrow\quad \mathbf{E}(z,t)=\eta\,\mathbf{H}(z,t)\times(-\hat{\mathbf{z}})$$

In the general case of Eq. (2.1.12), the E/H ratio does not remain constant. The Poynting vector and energy density consist of a part due to the forward wave and a part due to the backward one:

$$\begin{aligned} \boldsymbol{\mathcal{P}} &= \boldsymbol{E}\times\boldsymbol{H} = c\,\hat{\mathbf{z}}\,\big(\epsilon|\boldsymbol{F}|^2 - \epsilon|\boldsymbol{G}|^2\big) \\ w &= \frac{1}{2}\epsilon|\boldsymbol{E}|^2 + \frac{1}{2}\mu|\boldsymbol{H}|^2 = \epsilon|\boldsymbol{F}|^2 + \epsilon|\boldsymbol{G}|^2 \end{aligned} \tag{2.1.18}$$

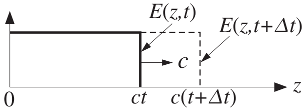
Example 2.1.1: A source located at $z = 0$ generates an electric field $\boldsymbol{E}(0,t)= \hat{\mathbf{x}}E_0\,u(t)$, where $u(t)$ is the unit-step function, and E_0, a constant. The field is launched towards the positive z-direction. Determine expressions for $\boldsymbol{E}(z,t)$ and $\boldsymbol{H}(z,t)$.

Solution: For a forward-moving wave, we have $\boldsymbol{E}(z,t)= \boldsymbol{F}(z-ct)= \boldsymbol{F}\big(0-c(t-z/c)\big)$, which implies that $\boldsymbol{E}(z,t)$ is completely determined by $\boldsymbol{E}(z,0)$, or alternatively, by $\boldsymbol{E}(0,t)$:

$$\boldsymbol{E}(z,t)= \boldsymbol{E}(z-ct,0)= \boldsymbol{E}(0,t-z/c)$$

Using this property, we find for the electric and magnetic fields:

$$\begin{aligned} \boldsymbol{E}(z,t) &= \boldsymbol{E}(0,t-z/c)= \hat{\mathbf{x}}E_0\,u(t-z/c) \\ \boldsymbol{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}(z,t)= \hat{\mathbf{y}}\,\frac{E_0}{\eta}\,u(t-z/c) \end{aligned}$$



Because of the unit-step, the non-zero values of the fields are restricted to $t-z/c \geq 0$, or, $z \leq ct$, that is, at time t the wavefront has propagated only up to position $z = ct$. The figure shows the expanding wavefronts at time t and $t+\Delta t$. □

Example 2.1.2: Consider the following three examples of electric fields specified at $t = 0$, and describing forward or backward fields as indicated:

$$\begin{aligned} &\boldsymbol{E}(z,0)= \hat{\mathbf{x}}E_0\cos(kz) && \text{(forward-moving)} \\ &\boldsymbol{E}(z,0)= \hat{\mathbf{y}}E_0\cos(kz) && \text{(backward-moving)} \\ &\boldsymbol{E}(z,0)= \hat{\mathbf{x}}E_1\cos(k_1z)+\hat{\mathbf{y}}E_2\cos(k_2z) && \text{(forward-moving)} \end{aligned}$$

where k, k_1, k_2 are given wavenumbers (measured in units of radians/m.) Determine the corresponding fields $\boldsymbol{E}(z,t)$ and $\boldsymbol{H}(z,t)$.

Solution: For the forward-moving cases, we replace z by $z-ct$, and for the backward-moving case, by $z+ct$. We find in the three cases:

$$\begin{aligned} \boldsymbol{E}(z,t) &= \hat{\mathbf{x}}E_0\cos\big(k(z-ct)\big) = \hat{\mathbf{x}}E_0\cos(\omega t-kz) \\ \boldsymbol{E}(z,t) &= \hat{\mathbf{y}}E_0\cos\big(k(z+ct)\big) = \hat{\mathbf{y}}E_0\cos(\omega t+kz) \\ \boldsymbol{E}(z,t) &= \hat{\mathbf{x}}E_1\cos(\omega_1 t-k_1z)+\hat{\mathbf{y}}E_2\cos(\omega_2 t-k_2z) \end{aligned}$$

where $\omega = kc$, and $\omega_1 = k_1c$, $\omega_2 = k_2c$. The corresponding magnetic fields are:

$$\begin{aligned} \boldsymbol{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}(z,t)= \hat{\mathbf{y}}\,\frac{E_0}{\eta}\,\cos(\omega t-kz) \qquad \text{(forward)} \\ \boldsymbol{H}(z,t) &= -\frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}(z,t)= \hat{\mathbf{x}}\,\frac{E_0}{\eta}\,\cos(\omega t+kz) \quad \text{(backward)} \\ \boldsymbol{H}(z,t) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}(z,t)= \hat{\mathbf{y}}\,\frac{E_1}{\eta}\,\cos(\omega_1 t-k_1z)-\hat{\mathbf{x}}\,\frac{E_2}{\eta}\,\cos(\omega_2 t-k_2z) \end{aligned}$$

The first two cases are single-frequency waves, and are discussed in more detail in the next section. The third case is a linear superposition of two waves with two different frequencies and polarizations. □

2.2 Monochromatic Waves

Uniform, single-frequency, plane waves propagating in a lossless medium are obtained as a special case of the previous section by assuming the harmonic time-dependence:

$$\begin{aligned} \boldsymbol{E}(x,y,z,t) &= \boldsymbol{E}(z)e^{j\omega t} \\ \boldsymbol{H}(x,y,z,t) &= \boldsymbol{H}(z)e^{j\omega t} \end{aligned} \tag{2.2.1}$$

where $\boldsymbol{E}(z)$ and $\boldsymbol{H}(z)$ are transverse with respect to the z-direction.

Maxwell's equations (2.1.5), or those of the decoupled system (2.1.10), may be solved very easily by replacing time derivatives by $\partial_t \to j\omega$. Then, Eqs. (2.1.10) become the first-order differential equations (see also Problem 2.3):

$$\frac{\partial\boldsymbol{E}_\pm(z)}{\partial z} = \mp jk\,\boldsymbol{E}_\pm(z)\,, \qquad \text{where} \quad k = \frac{\omega}{c} = \omega\sqrt{\mu\epsilon} \tag{2.2.2}$$

with solutions:

$$\begin{aligned} \boldsymbol{E}_+(z) &= \boldsymbol{E}_{0+}e^{-jkz} \qquad \text{(forward)} \\ \boldsymbol{E}_-(z) &= \boldsymbol{E}_{0-}e^{jkz} \qquad \text{(backward)} \end{aligned} \tag{2.2.3}$$

where $\boldsymbol{E}_{0\pm}$ are arbitrary (complex-valued) constant vectors such that $\hat{\mathbf{z}}\cdot\boldsymbol{E}_{0\pm} = 0$. The corresponding magnetic fields are:

$$\begin{aligned} \boldsymbol{H}_+(z) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}_+(z)= \frac{1}{\eta}\,(\hat{\mathbf{z}}\times\boldsymbol{E}_{0+})\,e^{-jkz} = \boldsymbol{H}_{0+}e^{-jkz} \\ \boldsymbol{H}_-(z) &= -\frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}_-(z)= -\frac{1}{\eta}\,(\hat{\mathbf{z}}\times\boldsymbol{E}_{0-})\,e^{jkz} = \boldsymbol{H}_{0-}e^{jkz} \end{aligned} \tag{2.2.4}$$

where we defined the constant amplitudes of the magnetic fields:

$$\boldsymbol{H}_{0\pm} = \pm\frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}_{0\pm} \tag{2.2.5}$$

Inserting (2.2.3) into (2.1.9), we obtain the general solution for single-frequency waves, expressed as a superposition of forward and backward components:

$$\boxed{\begin{aligned} \boldsymbol{E}(z) &= \boldsymbol{E}_{0+}e^{-jkz} + \boldsymbol{E}_{0-}e^{jkz} \\ \boldsymbol{H}(z) &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\big[\boldsymbol{E}_{0+}e^{-jkz} - \boldsymbol{E}_{0-}e^{jkz}\big] \end{aligned}} \qquad \text{(forward + backward waves)} \tag{2.2.6}$$

Setting $\boldsymbol{E}_{0\pm} = \hat{\mathbf{x}}A_\pm + \hat{\mathbf{y}}B_\pm$, and noting that $\hat{\mathbf{z}}\times\boldsymbol{E}_{0\pm} = \hat{\mathbf{z}}\times(\hat{\mathbf{x}}A_\pm + \hat{\mathbf{y}}B_\pm)= \hat{\mathbf{y}}A_\pm - \hat{\mathbf{x}}B_\pm$, we may rewrite (2.2.6) in terms of its cartesian components:

$$\begin{aligned} &E_x(z)= A_+e^{-jkz} + A_-e^{jkz}\,, && E_y(z)= B_+e^{-jkz} + B_-e^{jkz} \\ &H_y(z)= \frac{1}{\eta}\big[A_+e^{-jkz} - A_-e^{jkz}\big]\,, && H_x(z)= -\frac{1}{\eta}\big[B_+e^{-jkz} - B_-e^{jkz}\big] \end{aligned} \tag{2.2.7}$$

Wavefronts are defined, in general, to be the surfaces of constant phase. A forward moving wave $\boldsymbol{E}(z)=\boldsymbol{E}_0 e^{-jkz}$ corresponds to the time-varying field:

$$\boldsymbol{E}(z,t)=\boldsymbol{E}_0 e^{j\omega t-jkz}=\boldsymbol{E}_0 e^{-j\varphi(z,t)}\,, \qquad \text{where} \quad \varphi(z,t)=kz-\omega t$$

A surface of constant phase is obtained by setting $\varphi(z,t)=$ const. Denoting this constant by $\phi_0=kz_0$ and using the property $c=\omega/k$, we obtain the condition:

$$\varphi(z,t)=\varphi_0 \quad\Rightarrow\quad kz-\omega t=kz_0 \quad\Rightarrow\quad z=ct+z_0$$

Thus, the wavefront is the xy-plane intersecting the z-axis at the point $z=ct+z_0$, moving forward with velocity c. This justifies the term "plane wave."

A backward-moving wave will have planar wavefronts parametrized by $z=-ct+z_0$, that is, moving backwards. A wave that is a linear combination of forward and backward components, may be thought of as having two planar wavefronts, one moving forward, and the other backward.

The relationships (2.2.5) imply that the vectors $\{\boldsymbol{E}_{0+},\boldsymbol{H}_{0+},\hat{\mathbf{z}}\}$ and $\{\boldsymbol{E}_{0-},\boldsymbol{H}_{0-},-\hat{\mathbf{z}}\}$ will form right-handed orthogonal systems. The magnetic field $\boldsymbol{H}_{0\pm}$ is perpendicular to the electric field $\boldsymbol{E}_{0\pm}$ and the cross-product $\boldsymbol{E}_{0\pm}\times\boldsymbol{H}_{0\pm}$ points towards the direction of propagation, that is, $\pm\hat{\mathbf{z}}$. Fig. 2.2.1 depicts the case of a forward propagating wave.

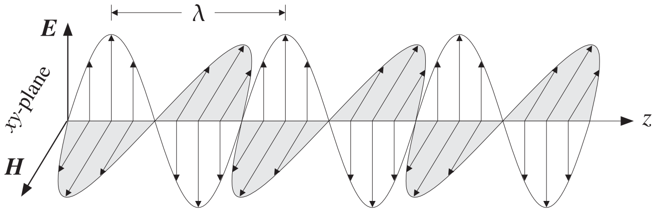


Fig. 2.2.1 Forward uniform plane wave.

The *wavelength* λ is the distance by which the phase of the sinusoidal wave changes by 2π radians. Since the propagation factor e^{-jkz} accumulates a phase of k radians per meter, we have by definition that $k\lambda=2\pi$. The wavelength λ can be expressed via the frequency of the wave in Hertz, $f=\omega/2\pi$, as follows:

$$\boxed{\lambda=\frac{2\pi}{k}=\frac{2\pi c}{\omega}=\frac{c}{f}} \tag{2.2.8}$$

If the propagation medium is free space, we use the vacuum values of the parameters $\{\epsilon,\mu,c,\eta\}$, that is, $\{\epsilon_0,\mu_0,c_0,\eta_0\}$. The free-space wavelength and corresponding wavenumber are:

$$\lambda_0=\frac{2\pi}{k_0}=\frac{c_0}{f}\,,\quad k_0=\frac{\omega}{c_0} \tag{2.2.9}$$

In a lossless but non-magnetic ($\mu=\mu_0$) dielectric with refractive index $n=\sqrt{\epsilon/\epsilon_0}$, the speed of light c, wavelength λ, and characteristic impedance η are all reduced by a scale factor n compared to the free-space values, whereas the wavenumber k is increased by a factor of n. Indeed, using the definitions $c=1/\sqrt{\mu_0\epsilon}$ and $\eta=\sqrt{\mu_0/\epsilon}$, we have:

$$\boxed{c=\frac{c_0}{n}\,,\quad \eta=\frac{\eta_0}{n}\,,\quad \lambda=\frac{\lambda_0}{n}\,,\quad k=nk_0} \tag{2.2.10}$$

Example 2.2.1: A microwave transmitter operating at the carrier frequency of 6 GHz is protected by a Plexiglas radome whose permittivity is $\epsilon=3\epsilon_0$.

The refractive index of the radome is $n=\sqrt{\epsilon/\epsilon_0}=\sqrt{3}=1.73$. The free-space wavelength and the wavelength inside the radome material are:

$$\lambda_0=\frac{c_0}{f}=\frac{3\times10^8}{6\times10^9}=0.05\text{ m}=5\text{ cm},\qquad \lambda=\frac{\lambda_0}{n}=\frac{5}{1.73}=2.9\text{ cm}$$

We will see later that if the radome is to be transparent to the wave, its thickness must be chosen to be equal to one-half wavelength, $l=\lambda/2$. Thus, $l=2.9/2=1.45$ cm. □

Example 2.2.2: The nominal speed of light in vacuum is $c_0=3\times10^8$ m/s. Because of the relationship $c_0=\lambda f$, it may be expressed in the following suggestive units that are appropriate in different application contexts:

$c_0 =$ 5000 km × 60 Hz	(power systems)
300 m × 1 MHz	(AM radio)
40 m × 7.5 MHz	(amateur radio)
3 m × 100 MHz	(FM radio, TV)
30 cm × 1 GHz	(cell phones)
10 cm × 3 GHz	(waveguides, radar)
3 cm × 10 GHz	(radar, satellites)
1.5 μm × 200 THz	(optical fibers)
500 nm × 600 THz	(visible spectrum)
100 nm × 3000 THz	(UV)

Similarly, in terms of length/time of propagation:

$c_0 =$ 36 000 km/120 msec	(geosynchronous satellites)
300 km/msec	(power lines)
300 m/μsec	(transmission lines)
30 cm/nsec	(circuit boards)

The typical half-wave monopole antenna (half of a half-wave dipole over a ground plane) has length $\lambda/4$ and is used in many applications, such as AM, FM, and cell phones. Thus, one can predict that the lengths of AM radio, FM radio, and cell phone antennas will be of the order of 75 m, 0.75 m, and 7.5 cm, respectively.

A more detailed list of electromagnetic frequency bands is given in Appendix B. The precise value of c_0 and the values of other physical constants are given in Appendix A. □

Wave propagation effects become important, and cannot be ignored, whenever the physical length of propagation is comparable to the wavelength λ. It follows from Eqs. (2.2.2) that the incremental change of a forward-moving electric field in propagating from z to $z+\Delta z$ is:

$$\frac{|\Delta\boldsymbol{E}_+|}{|\boldsymbol{E}_+|}=k\Delta z=2\pi\frac{\Delta z}{\lambda} \tag{2.2.11}$$

Thus, the change in the electric field can be ignored only if $\Delta z \ll \lambda$, otherwise, propagation effects must be taken into account.

For example, for an integrated circuit operating at 10 GHz, we have $\lambda = 3$ cm, which is comparable to the physical dimensions of the circuit.

Similarly, a cellular base station antenna is connected to the transmitter circuits by several meters of coaxial cable. For a 1-GHz system, the wavelength is 0.3 m, which implies that a 30-meter cable will be equivalent to 100 wavelengths.

2.3 Energy Density and Flux

The time-averaged energy density and flux of a uniform plane wave can be determined by Eq. (1.8.6). As in the previous section, the energy is shared equally by the electric and magnetic fields (in the forward or backward cases.) This is a general result for most wave propagation and waveguide problems.

The energy flux will be in the direction of propagation. For either a forward- or a backward-moving wave, we have from Eqs. (1.8.6) and (2.2.5):

$$w_e = \frac{1}{2}\,\mathrm{Re}\left[\frac{1}{2}\epsilon\,\mathbf{E}_\pm(z)\cdot\mathbf{E}_\pm^*(z)\right] = \frac{1}{2}\,\mathrm{Re}\left[\frac{1}{2}\epsilon\,\mathbf{E}_{0\pm}e^{-jkz}\cdot\mathbf{E}_{0\pm}^*e^{jkz}\right] = \frac{1}{4}\epsilon|\mathbf{E}_{0\pm}|^2$$

$$w_m = \frac{1}{2}\,\mathrm{Re}\left[\frac{1}{2}\mu\,\mathbf{H}_\pm(z)\cdot\mathbf{H}_\pm^*(z)\right] = \frac{1}{4}\mu|\mathbf{H}_{0\pm}|^2 = \frac{1}{4}\mu\frac{1}{\eta^2}|\hat{\mathbf{z}}\times\mathbf{E}_{0\pm}|^2 = \frac{1}{4}\epsilon|\mathbf{E}_{0\pm}|^2 = w_e$$

Thus, the electric and magnetic energy densities are equal and the total density is:

$$w = w_e + w_m = 2w_e = \frac{1}{2}\epsilon|\mathbf{E}_{0\pm}|^2 \tag{2.3.1}$$

For the time-averaged Poynting vector, we have similarly:

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\,\mathrm{Re}\big[\,\mathbf{E}_\pm(z)\times\mathbf{H}_\pm^*(z)\,\big] = \frac{1}{2\eta}\,\mathrm{Re}\big[\,\mathbf{E}_{0\pm}\times(\pm\hat{\mathbf{z}}\times\mathbf{E}_{0\pm}^*)\,\big]$$

Using the BAC-CAB rule and the orthogonality property $\hat{\mathbf{z}}\cdot\mathbf{E}_{0\pm} = 0$, we find:

$$\boldsymbol{\mathcal{P}} = \pm\hat{\mathbf{z}}\,\frac{1}{2\eta}|\mathbf{E}_{0\pm}|^2 = \pm c\,\hat{\mathbf{z}}\,\frac{1}{2}\,\epsilon|\mathbf{E}_{0\pm}|^2 \tag{2.3.2}$$

Thus, the energy flux is in the direction of propagation, that is, $\pm\hat{\mathbf{z}}$. The corresponding energy velocity is, as in the previous section:

$$\mathbf{v} = \frac{\boldsymbol{\mathcal{P}}}{w} = \pm c\,\hat{\mathbf{z}} \tag{2.3.3}$$

In the more general case of forward and backward waves, we find:

$$\begin{aligned} w &= \frac{1}{4}\,\mathrm{Re}\big[\epsilon\,\mathbf{E}(z)\cdot\mathbf{E}^*(z)+\mu\,\mathbf{H}(z)\cdot\mathbf{H}^*(z)\big] = \frac{1}{2}\epsilon|\mathbf{E}_{0+}|^2 + \frac{1}{2}\epsilon|\mathbf{E}_{0-}|^2 \\ \boldsymbol{\mathcal{P}} &= \frac{1}{2}\,\mathrm{Re}\big[\mathbf{E}(z)\times\mathbf{H}^*(z)\big] = \hat{\mathbf{z}}\left(\frac{1}{2\eta}|\mathbf{E}_{0+}|^2 - \frac{1}{2\eta}|\mathbf{E}_{0-}|^2\right) \end{aligned} \tag{2.3.4}$$

Thus, the total energy is the sum of the energies of the forward and backward components, whereas the net energy flux (to the right) is the difference between the forward and backward fluxes.

2.4 Wave Impedance

For forward or backward fields, the ratio of $\mathbf{E}(z)$ to $\mathbf{H}(z)\times\hat{\mathbf{z}}$ is constant and equal to the characteristic impedance of the medium. Indeed, it follows from Eq. (2.2.4) that

$$\mathbf{E}_\pm(z) = \pm\eta\mathbf{H}_\pm(z)\times\hat{\mathbf{z}}$$

However, this property is not true for the more general solution given by Eqs. (2.2.6). In general, the ratio of $\mathbf{E}(z)$ to $\mathbf{H}(z)\times\hat{\mathbf{z}}$ is called the *wave impedance*. Because of the vectorial character of the fields, we must define the ratio in terms of the corresponding x- and y-components:

$$\begin{aligned} Z_x(z) &= \frac{\big[\mathbf{E}(z)\big]_x}{\big[\mathbf{H}(z)\times\hat{\mathbf{z}}\big]_x} = \frac{E_x(z)}{H_y(z)} \\ Z_y(z) &= \frac{\big[\mathbf{E}(z)\big]_y}{\big[\mathbf{H}(z)\times\hat{\mathbf{z}}\big]_y} = -\frac{E_y(z)}{H_x(z)} \end{aligned} \qquad \text{(wave impedances)} \tag{2.4.1}$$

Using the cartesian expressions of Eq. (2.2.7), we find:

$$\begin{aligned} Z_x(z) &= \frac{E_x(z)}{H_y(z)} = \eta\,\frac{A_+e^{-jkz} + A_-e^{jkz}}{A_+e^{-jkz} - A_-e^{jkz}} \\ Z_y(z) &= -\frac{E_y(z)}{H_x(z)} = \eta\,\frac{B_+e^{-jkz} + B_-e^{jkz}}{B_+e^{-jkz} - B_-e^{jkz}} \end{aligned} \qquad \text{(wave impedances)} \tag{2.4.2}$$

Thus, the wave impedances are nontrivial functions of z. For forward waves (that is, with $A_- = B_- = 0$), we have $Z_x(z) = Z_y(z) = \eta$. For backward waves ($A_+ = B_+ = 0$), we have $Z_x(z) = Z_y(z) = -\eta$.

The wave impedance is a very useful concept in the subject of multiple dielectric interfaces and the matching of transmission lines. We will explore its use later on.

2.5 Polarization

Consider a forward-moving wave and let $\mathbf{E}_0 = \hat{\mathbf{x}}A_+ + \hat{\mathbf{y}}B_+$ be its complex-valued phasor amplitude, so that $\mathbf{E}(z) = \mathbf{E}_0e^{-jkz} = (\hat{\mathbf{x}}A_+ + \hat{\mathbf{y}}B_+)e^{-jkz}$. The time-varying field is obtained by restoring the factor $e^{j\omega t}$:

$$\mathbf{E}(z,t) = (\hat{\mathbf{x}}A_+ + \hat{\mathbf{y}}B_+)e^{j\omega t - jkz}$$

The polarization of a plane wave is defined to be the direction of the electric field. For example, if $B_+ = 0$, the E-field is along the x-direction and the wave will be linearly polarized.

More precisely, polarization is the direction of the time-varying real-valued field $\boldsymbol{\mathcal{E}}(z,t) = \mathrm{Re}\big[\mathbf{E}(z,t)\big]$. At any fixed point z, the vector $\boldsymbol{\mathcal{E}}(z,t)$ may be along a fixed linear direction or it may be rotating as a function of t, tracing a circle or an ellipse.

The polarization properties of the plane wave are determined by the relative magnitudes and phases of the complex-valued constants A_+, B_+. Writing them in their polar forms $A_+ = Ae^{j\phi_a}$ and $B_+ = Be^{j\phi_b}$, where A, B are *positive magnitudes*, we obtain:

$$\mathbf{E}(z,t) = \big(\hat{\mathbf{x}}Ae^{j\phi_a} + \hat{\mathbf{y}}Be^{j\phi_b}\big)e^{j\omega t - jkz} = \hat{\mathbf{x}}Ae^{j(\omega t - kz + \phi_a)} + \hat{\mathbf{y}}Be^{j(\omega t - kz + \phi_b)} \tag{2.5.1}$$

Extracting real parts and setting $\boldsymbol{\mathcal{E}}(z,t) = \mathrm{Re}\big[\mathbf{E}(z,t)\big] = \hat{\mathbf{x}}\mathcal{E}_x(z,t) + \hat{\mathbf{y}}\mathcal{E}_y(z,t)$, we find the corresponding real-valued x, y components:

$$\begin{aligned} \mathcal{E}_x(z,t) &= A\cos(\omega t - kz + \phi_a) \\ \mathcal{E}_y(z,t) &= B\cos(\omega t - kz + \phi_b) \end{aligned} \tag{2.5.2}$$

For a backward moving field, we replace k by $-k$ in the same expression. To determine the polarization of the wave, we consider the time-dependence of these fields at some *fixed* point along the z-axis, say at $z = 0$:

$$\begin{aligned} \mathcal{E}_x(t) &= A\cos(\omega t + \phi_a) \\ \mathcal{E}_y(t) &= B\cos(\omega t + \phi_b) \end{aligned} \tag{2.5.3}$$

The electric field vector $\boldsymbol{\mathcal{E}}(t) = \hat{\mathbf{x}}\mathcal{E}_x(t) + \hat{\mathbf{y}}\mathcal{E}_y(t)$ will be rotating on the xy-plane with angular frequency ω, with its tip tracing, in general, an ellipse. To see this, we expand Eq. (2.5.3) using a trigonometric identity:

$$\begin{aligned} \mathcal{E}_x(t) &= A\big[\cos\omega t\cos\phi_a - \sin\omega t\sin\phi_a\big] \\ \mathcal{E}_y(t) &= B\big[\cos\omega t\cos\phi_b - \sin\omega t\sin\phi_b\big] \end{aligned}$$

Solving for $\cos\omega t$ and $\sin\omega t$ in terms of $\mathcal{E}_x(t), \mathcal{E}_y(t)$, we find:

$$\begin{aligned} \cos\omega t\sin\phi &= \frac{\mathcal{E}_y(t)}{B}\sin\phi_a - \frac{\mathcal{E}_x(t)}{A}\sin\phi_b \\ \sin\omega t\sin\phi &= \frac{\mathcal{E}_y(t)}{B}\cos\phi_a - \frac{\mathcal{E}_x(t)}{A}\cos\phi_b \end{aligned}$$

where we defined the relative phase angle $\phi = \phi_a - \phi_b$.

Forming the sum of the squares of the two equations and using the trigonometric identity $\sin^2\omega t + \cos^2\omega t = 1$, we obtain a quadratic equation for the components $\mathcal{E}_x$ and $\mathcal{E}_y$, which describes an ellipse on the $\mathcal{E}_x, \mathcal{E}_y$ plane:

$$\left(\frac{\mathcal{E}_y(t)}{B}\sin\phi_a - \frac{\mathcal{E}_x(t)}{A}\sin\phi_b\right)^2 + \left(\frac{\mathcal{E}_y(t)}{B}\cos\phi_a - \frac{\mathcal{E}_x(t)}{A}\cos\phi_b\right)^2 = \sin^2\phi$$

This simplifies into:

$$\boxed{\frac{\mathcal{E}_x^2}{A^2} + \frac{\mathcal{E}_y^2}{B^2} - 2\cos\phi\,\frac{\mathcal{E}_x\mathcal{E}_y}{AB} = \sin^2\phi} \qquad \text{(polarization ellipse)} \tag{2.5.4}$$

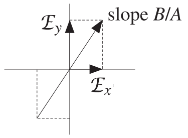
Depending on the values of the three quantities $\{A, B, \phi\}$ this *polarization ellipse* may be an ellipse, a circle, or a straight line. The electric field is accordingly called elliptically, circularly, or linearly polarized.

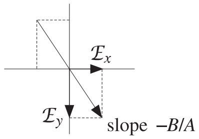
To get *linear polarization*, we set $\phi = 0$ or $\phi = \pi$, corresponding to $\phi_a = \phi_b = 0$, or $\phi_a = 0, \phi_b = -\pi$, so that the phasor amplitudes are $\mathbf{E}_0 = \hat{\mathbf{x}}A \pm \hat{\mathbf{y}}B$. Then, Eq. (2.5.4) degenerates into:

$$\frac{\mathcal{E}_x^2}{A^2} + \frac{\mathcal{E}_y^2}{B^2} \mp 2\,\frac{\mathcal{E}_x\mathcal{E}_y}{AB} = 0 \quad \Rightarrow \quad \left(\frac{\mathcal{E}_x}{A} \mp \frac{\mathcal{E}_y}{B}\right)^2 = 0$$

representing the straight lines:

$$\mathcal{E}_y = \pm\frac{B}{A}\,\mathcal{E}_x$$





The fields (2.5.2) take the forms, in the two cases $\phi = 0$ and $\phi = \pi$:

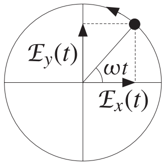
$$\begin{aligned} \mathcal{E}_x(t) &= A\cos\omega t \\ \mathcal{E}_y(t) &= B\cos\omega t \end{aligned} \quad \text{and} \quad \begin{aligned} \mathcal{E}_x(t) &= A\cos\omega t \\ \mathcal{E}_y(t) &= B\cos(\omega t - \pi) = -B\cos\omega t \end{aligned}$$

To get *circular polarization*, we set $A = B$ and $\phi = \pm\pi/2$. In this case, the polarization ellipse becomes the equation of a circle:

$$\frac{\mathcal{E}_x^2}{A^2} + \frac{\mathcal{E}_y^2}{A^2} = 1$$

The *sense* of rotation, in conjunction with the direction of propagation, defines left-circular versus right-circular polarization. For the case, $\phi_a = 0$ and $\phi_b = -\pi/2$, we have $\phi = \phi_a - \phi_b = \pi/2$ and complex amplitude $\mathbf{E}_0 = A(\hat{\mathbf{x}} - j\hat{\mathbf{y}})$. Then,

$$\begin{aligned} \mathcal{E}_x(t) &= A\cos\omega t \\ \mathcal{E}_y(t) &= A\cos(\omega t - \pi/2) = A\sin\omega t \end{aligned}$$



Thus, the tip of the electric field vector rotates *counterclockwise* on the xy-plane. To decide whether this represents right or left circular polarization, we use the *IEEE convention* [92], which is as follows.

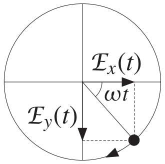
Curl the fingers of your left and right hands into a fist and point both thumbs towards the direction of propagation. If the fingers of your right (left) hand are curling in the direction of rotation of the electric field, then the polarization is right (left) polarized.[†]

Thus, in the present example, because we had a forward-moving field and the field is turning counterclockwise, the polarization will be right-circular. If the field were moving backwards, then it would be left-circular. For the case, $\phi = -\pi/2$, arising from $\phi_a = 0$

[†] Most engineering texts use the IEEE convention and most physics texts, the opposite convention.

and $\phi_b = \pi/2$, we have complex amplitude $\mathbf{E}_0 = A\,(\hat{\mathbf{x}} + j\hat{\mathbf{y}})$. Then, Eq. (2.5.3) becomes:

$$\mathcal{E}_x(t) = A\cos\omega t$$
$$\mathcal{E}_y(t) = A\cos(\omega t + \pi/2) = -A\sin\omega t$$



The tip of the electric field vector rotates *clockwise* on the xy-plane. Since the wave is moving forward, this will represent left-circular polarization. Fig. 2.5.1 depicts the four cases of left/right polarization with forward/backward waves.

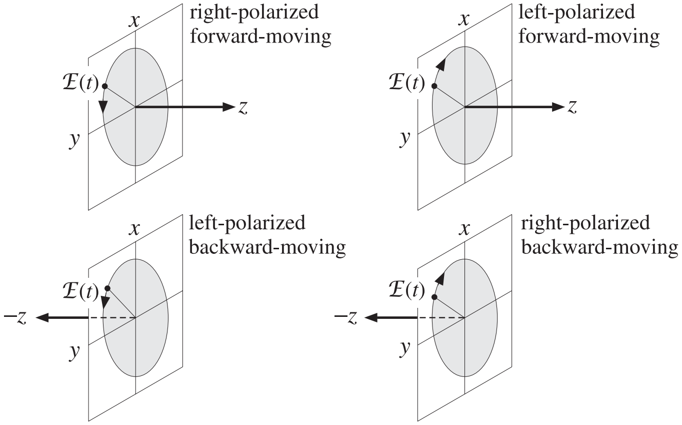


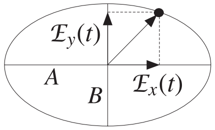
Fig. 2.5.1 Left and right circular polarizations.

To summarize, the electric field of a *circularly polarized* uniform plane wave will be, in its phasor form:

$\mathbf{E}(z) = A\,(\hat{\mathbf{x}} - j\hat{\mathbf{y}})\,e^{-jkz}$ (right-polarized, forward-moving)
$\mathbf{E}(z) = A\,(\hat{\mathbf{x}} + j\hat{\mathbf{y}})\,e^{-jkz}$ (left-polarized, forward-moving)
$\mathbf{E}(z) = A\,(\hat{\mathbf{x}} - j\hat{\mathbf{y}})\,e^{jkz}$ (left-polarized, backward-moving)
$\mathbf{E}(z) = A\,(\hat{\mathbf{x}} + j\hat{\mathbf{y}})\,e^{jkz}$ (right-polarized, backward-moving)

If $A \neq B$, but the phase difference is still $\phi = \pm\pi/2$, we get an ellipse with major and minor axes oriented along the x, y directions. Eq. (2.5.4) will be now:

$$\frac{\mathcal{E}_x^2}{A^2} + \frac{\mathcal{E}_y^2}{B^2} = 1$$



Finally, if $A \neq B$ and ϕ is arbitrary, then the major/minor axes of the ellipse (2.5.4) will be rotated relative to the x, y directions. Fig. 2.5.2 illustrates the general case.

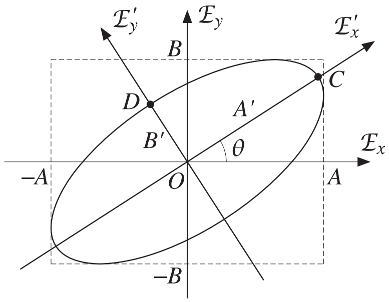


Fig. 2.5.2 General polarization ellipse.

It can be shown (see Problem 2.15) that the tilt angle θ is given by:

$$\tan 2\theta = \frac{2AB}{A^2 - B^2}\cos\phi \tag{2.5.5}$$

The ellipse semi-axes A', B', that is, the lengths OC and OD, are given by:

$$\begin{aligned} A' &= \sqrt{\frac{1}{2}(A^2 + B^2) + \frac{s}{2}\sqrt{(A^2 - B^2)^2 + 4A^2B^2\cos^2\phi}} \\ B' &= \sqrt{\frac{1}{2}(A^2 + B^2) - \frac{s}{2}\sqrt{(A^2 - B^2)^2 + 4A^2B^2\cos^2\phi}} \end{aligned} \tag{2.5.6}$$

where $s = \mathrm{sign}(A - B)$. These results are obtained by defining the rotated coordinate system of the ellipse axes:

$$\begin{aligned} \mathcal{E}'_x &= \mathcal{E}_x\cos\theta + \mathcal{E}_y\sin\theta \\ \mathcal{E}'_y &= \mathcal{E}_y\cos\theta - \mathcal{E}_x\sin\theta \end{aligned} \tag{2.5.7}$$

and showing that Eq. (2.5.4) transforms into the standardized form:

$$\frac{\mathcal{E}'^2_x}{A'^2} + \frac{\mathcal{E}'^2_y}{B'^2} = 1 \tag{2.5.8}$$

The polarization ellipse is bounded by the rectangle with sides at the end-points $\pm A, \pm B$, as shown in the figure. To decide whether the elliptic polarization is left- or right-handed, we may use the same rules depicted in Fig. 2.5.1

Example 2.5.1: Determine the real-valued electric and magnetic field components and the polarization of the following fields specified in their phasor form (given in units of V/m):

a. $\mathbf{E}(z) = -3j\,\hat{\mathbf{x}}\,e^{-jkz}$
b. $\mathbf{E}(z) = \big(3\,\hat{\mathbf{x}} + 4\,\hat{\mathbf{y}}\big)e^{+jkz}$
c. $\mathbf{E}(z) = \big(-4\,\hat{\mathbf{x}} + 3\,\hat{\mathbf{y}}\big)e^{-jkz}$
d. $\mathbf{E}(z) = \big(3e^{j\pi/3}\,\hat{\mathbf{x}} + 3\,\hat{\mathbf{y}}\big)e^{+jkz}$
e. $\mathbf{E}(z) = \big(4\,\hat{\mathbf{x}} + 3e^{-j\pi/4}\,\hat{\mathbf{y}}\big)e^{-jkz}$
f. $\mathbf{E}(z) = \big(3e^{-j\pi/8}\,\hat{\mathbf{x}} + 4e^{j\pi/8}\,\hat{\mathbf{y}}\big)e^{+jkz}$
g. $\mathbf{E}(z) = \big(4e^{j\pi/4}\,\hat{\mathbf{x}} + 3e^{-j\pi/2}\,\hat{\mathbf{y}}\big)e^{-jkz}$
h. $\mathbf{E}(z) = \big(3e^{-j\pi/2}\,\hat{\mathbf{x}} + 4e^{j\pi/4}\,\hat{\mathbf{y}}\big)e^{+jkz}$

Solution: Restoring the $e^{j\omega t}$ factor and taking real-parts, we find the x, y electric field components, according to Eq. (2.5.2):

$$
\begin{array}{lll}
\text{a.} & \mathcal{E}_x(z,t) = 3\cos(\omega t - kz - \pi/2)\,, & \mathcal{E}_y(z,t) = 0 \\
\text{b.} & \mathcal{E}_x(z,t) = 3\cos(\omega t + kz)\,, & \mathcal{E}_y(z,t) = 4\cos(\omega t + kz) \\
\text{c.} & \mathcal{E}_x(z,t) = 4\cos(\omega t - kz + \pi)\,, & \mathcal{E}_y(z,t) = 3\cos(\omega t - kz) \\
\text{d.} & \mathcal{E}_x(z,t) = 3\cos(\omega t + kz + \pi/3)\,, & \mathcal{E}_y(z,t) = 3\cos(\omega t + kz) \\
\text{e.} & \mathcal{E}_x(z,t) = 4\cos(\omega t - kz)\,, & \mathcal{E}_y(z,t) = 3\cos(\omega t - kz - \pi/4) \\
\text{f.} & \mathcal{E}_x(z,t) = 3\cos(\omega t + kz - \pi/8)\,, & \mathcal{E}_y(z,t) = 4\cos(\omega t + kz + \pi/8) \\
\text{g.} & \mathcal{E}_x(z,t) = 4\cos(\omega t - kz + \pi/4)\,, & \mathcal{E}_y(z,t) = 3\cos(\omega t - kz - \pi/2) \\
\text{h.} & \mathcal{E}_x(z,t) = 3\cos(\omega t + kz - \pi/2)\,, & \mathcal{E}_y(z,t) = 4\cos(\omega t + kz + \pi/4)
\end{array}
$$

Since these are either forward or backward waves, the corresponding magnetic fields are obtained by using the formula $\boldsymbol{\mathcal{H}}(z,t) = \pm\hat{\mathbf{z}}\times\boldsymbol{\mathcal{E}}(z,t)/\eta$. This gives the x, y components:

$$
\begin{array}{lll}
\text{(cases a, c, e, g):} & \mathcal{H}_x(z,t) = -\dfrac{1}{\eta}\mathcal{E}_y(z,t)\,, & \mathcal{H}_y(z,t) = \dfrac{1}{\eta}\mathcal{E}_x(z,t) \\
\text{(cases b, d, f, h):} & \mathcal{H}_x(z,t) = \dfrac{1}{\eta}\mathcal{E}_y(z,t)\,, & \mathcal{H}_y(z,t) = -\dfrac{1}{\eta}\mathcal{E}_x(z,t)
\end{array}
$$

To determine the polarization vectors, we evaluate the electric fields at $z = 0$:

$$
\begin{array}{lll}
\text{a.} & \mathcal{E}_x(t) = 3\cos(\omega t - \pi/2)\,, & \mathcal{E}_y(t) = 0 \\
\text{b.} & \mathcal{E}_x(t) = 3\cos(\omega t)\,, & \mathcal{E}_y(t) = 4\cos(\omega t) \\
\text{c.} & \mathcal{E}_x(t) = 4\cos(\omega t + \pi)\,, & \mathcal{E}_y(t) = 3\cos(\omega t) \\
\text{d.} & \mathcal{E}_x(t) = 3\cos(\omega t + \pi/3)\,, & \mathcal{E}_y(t) = 3\cos(\omega t) \\
\text{e.} & \mathcal{E}_x(t) = 4\cos(\omega t)\,, & \mathcal{E}_y(t) = 3\cos(\omega t - \pi/4) \\
\text{f.} & \mathcal{E}_x(t) = 3\cos(\omega t - \pi/8)\,, & \mathcal{E}_y(t) = 4\cos(\omega t + \pi/8) \\
\text{g.} & \mathcal{E}_x(t) = 4\cos(\omega t + \pi/4)\,, & \mathcal{E}_y(t) = 3\cos(\omega t - \pi/2) \\
\text{h.} & \mathcal{E}_x(t) = 3\cos(\omega t - \pi/2)\,, & \mathcal{E}_y(t) = 4\cos(\omega t + \pi/4)
\end{array}
$$

The polarization ellipse parameters A, B, and $\phi = \phi_a - \phi_b$, as well as the computed semi-major axes A', B', tilt angle θ, sense of rotation of the electric field, and polarization type are given below:

case	A	B	ϕ	A'	B'	θ	rotation	polarization
a.	3	0	-90^{o}	3	0	0^{o}	→	linear/forward
b.	3	4	0^{o}	0	5	-36.87^{o}	↗	linear/backward
c.	4	3	180^{o}	5	0	-36.87^{o}	↖	linear/forward
d.	3	3	60^{o}	3.674	2.121	45^{o}	↺	left/backward
e.	4	3	45^{o}	4.656	1.822	33.79^{o}	↺	right/forward
f.	3	4	-45^{o}	1.822	4.656	-33.79^{o}	↻	right/backward
g.	4	3	135^{o}	4.656	1.822	-33.79^{o}	↺	right/forward
h.	3	4	-135^{o}	1.822	4.656	33.79^{o}	↻	right/backward

In the linear case (b), the polarization ellipse collapses along its A'-axis ($A' = 0$) and becomes a straight line along its B'-axis. The tilt angle θ still measures the angle of the A'-axis from the x-axis. The actual direction of the electric field will be $90^{\text{o}} - 36.87^{\text{o}} = 53.13^{\text{o}}$, which is equal to the slope angle, $\text{atan}(B/A) = \text{atan}(4/3) = 53.13^{\text{o}}$.

In case (c), the ellipse collapses along its B'-axis. Therefore, θ coincides with the angle of the slope of the electric field vector, that is, $\text{atan}(-B/A) = \text{atan}(-3/4) = -36.87^{\text{o}}$. □

With the understanding that θ always represents the slope of the A'-axis (whether collapsed or not, major or minor), Eqs. (2.5.5) and (2.5.6) correctly calculate all the special cases, except when $A = B$, which has tilt angle and semi-axes:

$$\theta = 45^{\text{o}}\,, \quad A' = A\sqrt{1+\cos\phi}\,, \quad B' = A\sqrt{1-\cos\phi} \tag{2.5.9}$$

The MATLAB function `ellipse.m` calculates the ellipse semi-axes and tilt angle, A', B', θ, given the parameters A, B, ϕ. It has usage:

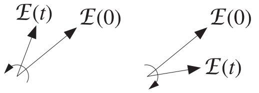
```
[a,b,th] = ellipse(A,B,phi)       % polarization ellipse parameters
```

For example, the function will return the values of the A', B', θ columns of the previous example, if it is called with the inputs:

```
A   = [3, 3, 4, 3, 4, 3, 4, 3]';
B   = [0, 4, 3, 3, 3, 4, 3, 4]';
phi = [-90, 0, 180, 60, 45, -45, 135, -135]';
```

To determine quickly the sense of rotation around the polarization ellipse, we use the rule that the rotation will be counterclockwise if the phase difference $\phi = \phi_a - \phi_b$ is such that $\sin\phi > 0$, and clockwise, if $\sin\phi < 0$. This can be seen by considering the electric field at time $t = 0$ and at a neighboring time t. Using Eq. (2.5.3), we have:

$$
\begin{aligned}
\boldsymbol{\mathcal{E}}(0) &= \hat{\mathbf{x}}A\cos\phi_a + \hat{\mathbf{y}}B\cos\phi_b \\
\boldsymbol{\mathcal{E}}(t) &= \hat{\mathbf{x}}A\cos(\omega t + \phi_a) + \hat{\mathbf{y}}B\cos(\omega t + \phi_b)
\end{aligned}
$$



The sense of rotation may be determined from the cross-product $\boldsymbol{\mathcal{E}}(0)\times\boldsymbol{\mathcal{E}}(t)$. If the rotation is counterclockwise, this vector will point towards the positive z-direction, and otherwise, it will point towards the negative z-direction. It follows easily that:

$$\boldsymbol{\mathcal{E}}(0)\times\boldsymbol{\mathcal{E}}(t) = \hat{\mathbf{z}}AB\sin\phi\sin\omega t \tag{2.5.10}$$

Thus, for t small and positive (such that $\sin\omega t > 0$), the direction of the vector $\boldsymbol{\mathcal{E}}(0)\times\boldsymbol{\mathcal{E}}(t)$ is determined by the sign of $\sin\phi$.

2.6 Uniform Plane Waves in Lossy Media

We saw in Sec. 1.9 that power losses may arise because of conduction and/or material polarization. A wave propagating in a lossy medium will set up a conduction current $\boldsymbol{J}_{\text{cond}} = \sigma\boldsymbol{E}$ and a displacement (polarization) current $\boldsymbol{J}_{\text{disp}} = j\omega\boldsymbol{D} = j\omega\epsilon_d\boldsymbol{E}$. Both currents will cause ohmic losses. The total current is the sum:

$$\boldsymbol{J}_{\text{tot}} = \boldsymbol{J}_{\text{cond}} + \boldsymbol{J}_{\text{disp}} = (\sigma + j\omega\epsilon_d)\boldsymbol{E} = j\omega\epsilon_c\boldsymbol{E}$$

where ϵ_c is the effective complex dielectric constant introduced in Eq. (1.9.29):

$$j\omega\epsilon_c = \sigma + j\omega\epsilon_d \quad\Rightarrow\quad \epsilon_c = \epsilon_d - j\,\frac{\sigma}{\omega} \tag{2.6.1}$$

The quantities σ, ϵ_d may be complex-valued and frequency-dependent. However, we will assume that over the desired frequency band of interest, the conductivity σ is real-valued; the permittivity of the dielectric may be assumed to be complex, $\epsilon_d = \epsilon_d' - j\epsilon_d''$. Thus, the effective ϵ_c has real and imaginary parts:

$$\boxed{\epsilon_c = \epsilon' - j\epsilon'' = \epsilon_d' - j\left(\epsilon_d'' + \frac{\sigma}{\omega}\right)} \tag{2.6.2}$$

Power losses arise from the non-zero imaginary part ϵ''. We recall from Eq. (1.9.32) that the time-averaged ohmic power losses per unit volume are given by:

$$\frac{dP_{\text{loss}}}{dV} = \frac{1}{2}\,\text{Re}\big[\mathbf{J}_{\text{tot}}\cdot\mathbf{E}^*\big] = \frac{1}{2}\,\omega\epsilon''\,|\mathbf{E}|^2 = \frac{1}{2}\,(\sigma + \omega\epsilon_d'')\,|\mathbf{E}|^2 \tag{2.6.3}$$

Uniform plane waves propagating in such lossy medium will satisfy Maxwell's equations (1.8.2), with the right-hand side of Ampère's law given by $\mathbf{J}_{\text{tot}} = \mathbf{J} + j\omega\mathbf{D} = j\omega\epsilon_c\mathbf{E}$.

The assumption of uniformity ($\partial_x = \partial_y = 0$), will imply again that the fields $\mathbf{E}, \mathbf{H}$ are transverse to the direction $\hat{\mathbf{z}}$. Then, Faraday's and Ampère's equations become:

$$\begin{aligned} \nabla\times\mathbf{E} &= -j\omega\mu\mathbf{H} \\ \nabla\times\mathbf{H} &= j\omega\epsilon_c\mathbf{E} \end{aligned} \quad\Rightarrow\quad \begin{aligned} \hat{\mathbf{z}}\times\partial_z\mathbf{E} &= -j\omega\mu\mathbf{H} \\ \hat{\mathbf{z}}\times\partial_z\mathbf{H} &= j\omega\epsilon_c\mathbf{E} \end{aligned} \tag{2.6.4}$$

These may be written in a more convenient form by introducing the complex wavenumber k_c and complex characteristic impedance η_c defined by:

$$\boxed{k_c = \omega\sqrt{\mu\epsilon_c}\,, \quad \eta_c = \sqrt{\frac{\mu}{\epsilon_c}}} \tag{2.6.5}$$

They correspond to the usual definitions $k = \omega/c = \omega\sqrt{\mu\epsilon}$ and $\eta = \sqrt{\mu/\epsilon}$ with the replacement $\epsilon \to \epsilon_c$. Noting that $\omega\mu = k_c\eta_c$ and $\omega\epsilon_c = k_c/\eta_c$, Eqs. (2.6.4) may be written in the following form (using the orthogonality property $\hat{\mathbf{z}}\cdot\mathbf{E} = 0$ and the BAC-CAB rule on the first equation):

$$\frac{\partial}{\partial z}\begin{bmatrix} \mathbf{E} \\ \eta_c\mathbf{H}\times\hat{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} 0 & -jk_c \\ -jk_c & 0 \end{bmatrix}\begin{bmatrix} \mathbf{E} \\ \eta_c\mathbf{H}\times\hat{\mathbf{z}} \end{bmatrix} \tag{2.6.6}$$

To decouple them, we introduce the forward and backward electric fields:

$$\begin{aligned} \mathbf{E}_+ &= \frac{1}{2}\,(\mathbf{E} + \eta_c\mathbf{H}\times\hat{\mathbf{z}}) \\ \mathbf{E}_- &= \frac{1}{2}\,(\mathbf{E} - \eta_c\mathbf{H}\times\hat{\mathbf{z}}) \end{aligned} \quad\Leftrightarrow\quad \begin{aligned} \mathbf{E} &= \mathbf{E}_+ + \mathbf{E}_- \\ \mathbf{H} &= \frac{1}{\eta_c}\,\hat{\mathbf{z}}\times[\mathbf{E}_+ - \mathbf{E}_-] \end{aligned} \tag{2.6.7}$$

Then, Eqs. (2.6.6) may be replaced by the equivalent system:

$$\frac{\partial}{\partial z}\begin{bmatrix} \mathbf{E}_+ \\ \mathbf{E}_- \end{bmatrix} = \begin{bmatrix} -jk_c & 0 \\ 0 & jk_c \end{bmatrix}\begin{bmatrix} \mathbf{E}_+ \\ \mathbf{E}_- \end{bmatrix} \tag{2.6.8}$$

with solutions:

$$\mathbf{E}_\pm(z) = \mathbf{E}_{0\pm}e^{\mp jk_cz}\,, \qquad \text{where}\quad \hat{\mathbf{z}}\cdot\mathbf{E}_{0\pm} = 0 \tag{2.6.9}$$

Thus, the propagating electric and magnetic fields are linear combinations of forward and backward components:

$$\boxed{\begin{aligned} \mathbf{E}(z) &= \mathbf{E}_{0+}e^{-jk_cz} + \mathbf{E}_{0-}e^{jk_cz} \\ \mathbf{H}(z) &= \frac{1}{\eta_c}\,\hat{\mathbf{z}}\times\big[\mathbf{E}_{0+}e^{-jk_cz} - \mathbf{E}_{0-}e^{jk_cz}\big] \end{aligned}} \tag{2.6.10}$$

In particular, for a forward-moving wave we have:

$$\mathbf{E}(z) = \mathbf{E}_0e^{-jk_cz}\,, \quad \mathbf{H}(z) = \mathbf{H}_0e^{-jk_cz}\,, \quad \text{with}\quad \hat{\mathbf{z}}\cdot\mathbf{E}_0 = 0\,, \quad \mathbf{H}_0 = \frac{1}{\eta_c}\,\hat{\mathbf{z}}\times\mathbf{E}_0 \tag{2.6.11}$$

Eqs. (2.6.10) are the same as in the lossless case but with the replacements $k \to k_c$ and $\eta \to \eta_c$. The lossless case is obtained in the limit of a purely real-valued ϵ_c.

Because k_c is complex-valued, we define the *phase* and *attenuation* constants β and α as the real and imaginary parts of k_c, that is,

$$k_c = \beta - j\alpha = \omega\sqrt{\mu(\epsilon' - j\epsilon'')} \tag{2.6.12}$$

We may also define a *complex refractive index* $n_c = k_c/k_0$ that measures k_c relative to its free-space value $k_0 = \omega/c_0 = \omega\sqrt{\mu_0\epsilon_0}$. For a non-magnetic medium, we have:

$$n_c = \frac{k_c}{k_0} = \sqrt{\frac{\epsilon_c}{\epsilon_0}} = \sqrt{\frac{\epsilon' - j\epsilon''}{\epsilon_0}} \equiv n_r - jn_i \tag{2.6.13}$$

where n_r, n_i are the real and imaginary parts of n_c. The quantity n_i is called the *extinction coefficient* and n_r, the refractive index. Another commonly used notation is the *propagation constant* γ defined by:

$$\gamma = jk_c = \alpha + j\beta \tag{2.6.14}$$

It follows from $\gamma = \alpha + j\beta = jk_c = jk_0n_c = jk_0(n_r - jn_i)$ that $\beta = k_0n_r$ and $\alpha = k_0n_i$. The nomenclature about phase and attenuation constants has its origins in the propagation factor e^{-jk_cz}. We can write it in the alternative forms:

$$\boxed{e^{-jk_cz} = e^{-\gamma z} = e^{-\alpha z}e^{-j\beta z} = e^{-k_0n_iz}e^{-jk_0n_rz}} \tag{2.6.15}$$

Thus, the wave amplitudes *attenuate exponentially* with the factor $e^{-\alpha z}$, and oscillate with the phase factor $e^{-j\beta z}$. The energy of the wave attenuates by the factor $e^{-2\alpha z}$, as can be seen by computing the Poynting vector. Because e^{-jk_cz} is no longer a pure phase factor and η_c is not real, we have for the forward-moving wave of Eq. (2.6.11):

$$\begin{aligned} \boldsymbol{\mathcal{P}}(z) &= \frac{1}{2}\,\text{Re}\big[\mathbf{E}(z)\times\mathbf{H}^*(z)\big] = \frac{1}{2}\,\text{Re}\left[\frac{1}{\eta_c^*}\,\mathbf{E}_0\times(\hat{\mathbf{z}}\times\mathbf{E}_0^*)\,e^{-(\alpha+j\beta)z}e^{-(\alpha-j\beta)z}\right] \\ &= \hat{\mathbf{z}}\,\frac{1}{2}\,\text{Re}(\eta_c^{-1})\,|\mathbf{E}_0|^2e^{-2\alpha z} = \hat{\mathbf{z}}\,\mathcal{P}(0)\,e^{-2\alpha z} = \hat{\mathbf{z}}\,\mathcal{P}(z) \end{aligned}$$

Thus, the power per unit area flowing past the point z in the forward z-direction will be:

$$\boxed{\mathcal{P}(z)=\mathcal{P}(0)\,e^{-2\alpha z}} \tag{2.6.16}$$

The quantity $\mathcal{P}(0)$ is the power per unit area flowing past the point $z=0$. Denoting the real and imaginary parts of η_c by η',η'', so that, $\eta_c=\eta'+j\eta''$, and noting that $|\boldsymbol{E}_0|=|\eta_c\boldsymbol{H}_0\times\hat{\mathbf{z}}|=|\eta_c||\boldsymbol{H}_0|$, we may express $\mathcal{P}(0)$ in the equivalent forms:

$$\mathcal{P}(0)=\frac{1}{2}\,\mathrm{Re}\big(\eta_c^{-1}\big)\,|\boldsymbol{E}_0|^2=\frac{1}{2}\eta'\,|\boldsymbol{H}_0|^2 \tag{2.6.17}$$

The attenuation coefficient α is measured in *nepers per meter*. However, a more practical way of expressing the power attenuation is in dB per meter. Taking logs of Eq. (2.6.16), we have for the dB attenuation at z, relative to $z=0$:

$$A_{\mathrm{dB}}(z)=-10\log_{10}\left[\frac{\mathcal{P}(z)}{\mathcal{P}(0)}\right]=20\log_{10}(e)\,\alpha z=8.686\,\alpha z \tag{2.6.18}$$

where we used the numerical value $20\log_{10}e=8.686$. Thus, the quantity $\alpha_{\mathrm{dB}}=8.686\,\alpha$ is the attenuation in *dB per meter*:

$$\alpha_{\mathrm{dB}}=8.686\,\alpha \qquad (\mathrm{dB/m}) \tag{2.6.19}$$

Another way of expressing the power attenuation is by means of the so-called *penetration* or *skin depth* defined as the inverse of α:

$$\boxed{\delta=\frac{1}{\alpha}} \qquad \text{(skin depth)} \tag{2.6.20}$$

Then, Eq. (2.6.18) can be rewritten in the form:

$$A_{\mathrm{dB}}(z)=8.686\,\frac{z}{\delta} \qquad \text{(attenuation in dB)} \tag{2.6.21}$$

This gives rise to the so-called “9-dB per delta” rule, that is, every time z is increased by a distance δ, the attenuation increases by $8.686\simeq 9$ dB.

A useful way to represent Eq. (2.6.16) in practice is to consider its infinitesimal version obtained by differentiating it with respect to z and solving for α:

$$\mathcal{P}'(z)=-2\alpha\mathcal{P}(0)\,e^{-2\alpha z}=2\alpha\mathcal{P}(z) \quad\Rightarrow\quad \alpha=-\frac{\mathcal{P}'(z)}{2\mathcal{P}(z)}$$

The quantity $\mathcal{P}'_{\mathrm{loss}}=-\mathcal{P}'$ represents the power lost from the wave per unit length (in the propagation direction.) Thus, the attenuation coefficient is the ratio of the power loss per unit length to twice the power transmitted:

$$\boxed{\alpha=\frac{P'_{\mathrm{loss}}}{2P_{\mathrm{transm}}}} \qquad \text{(attenuation coefficient)} \tag{2.6.22}$$

If there are several physical mechanisms for the power loss, then α becomes the *sum* over all possible cases. For example, in a waveguide or a coaxial cable filled with a slightly lossy dielectric, power will be lost because of the small conduction/polarization currents set up within the dielectric and also because of the ohmic losses in the walls of the guiding conductors, so that the total α will be $\alpha=\alpha_{\mathrm{diel}}+\alpha_{\mathrm{walls}}$.

Next, we verify that the exponential loss of power from the propagating wave is due to ohmic heat losses. In Fig. 2.6.1, we consider a volume $dV=l\,dA$ of area dA and length l along the z-direction.

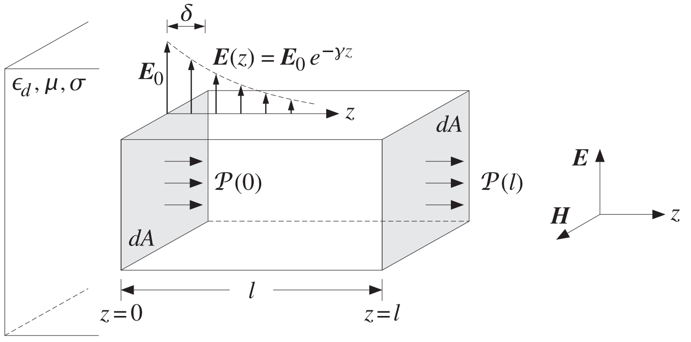


Fig. 2.6.1 Power flow in lossy dielectric.

From the definition of $\mathcal{P}(z)$ as power flow per unit area, it follows that the power entering the area dA at $z=0$ will be $dP_{\mathrm{in}}=\mathcal{P}(0)\,dA$, and the power leaving that area at $z=l$, $dP_{\mathrm{out}}=\mathcal{P}(l)\,dA$. The difference $dP_{\mathrm{loss}}=dP_{\mathrm{in}}-dP_{\mathrm{out}}=\big[\mathcal{P}(0)-\mathcal{P}(l)\big]dA$ will be the power lost from the wave within the volume $l\,dA$. Because $\mathcal{P}(l)=\mathcal{P}(0)\,e^{-2\alpha l}$, we have for the power loss per unit area:

$$\frac{dP_{\mathrm{loss}}}{dA}=\mathcal{P}(0)-\mathcal{P}(l)=\mathcal{P}(0)\big(1-e^{-2\alpha l}\big)=\frac{1}{2}\,\mathrm{Re}\big(\eta_c^{-1}\big)\,|\boldsymbol{E}_0|^2\big(1-e^{-2\alpha l}\big) \tag{2.6.23}$$

On the other hand, according to Eq. (2.6.3), the ohmic power loss per unit volume will be $\omega\epsilon''|\boldsymbol{E}(z)|^2/2$. Integrating this quantity from $z=0$ to $z=l$ will give the total ohmic losses within the volume $l\,dA$ of Fig. 2.6.1. Thus, we have:

$$dP_{\mathrm{ohmic}}=\frac{1}{2}\omega\epsilon''\int_0^l|\boldsymbol{E}(z)|^2\,dz\,dA=\frac{1}{2}\omega\epsilon''\left[\int_0^l|\boldsymbol{E}_0|^2e^{-2\alpha z}\,dz\right]dA\,, \qquad \text{or,}$$

$$\frac{dP_{\mathrm{ohmic}}}{dA}=\frac{\omega\epsilon''}{4\alpha}\,|\boldsymbol{E}_0|^2\big(1-e^{-2\alpha l}\big) \tag{2.6.24}$$

Are the two expressions in Eqs. (2.6.23) and (2.6.24) equal? The answer is yes, as follows from the relationship among the quantities η_c,ϵ'',α (see Problem 2.17):

$$\boxed{\mathrm{Re}\big(\eta_c^{-1}\big)=\frac{\omega\epsilon''}{2\alpha}} \tag{2.6.25}$$

Thus, the power lost from the wave is entirely accounted for by the ohmic losses within the propagation medium. The equality of (2.6.23) and (2.6.24) is an example of the more general relationship proved in Problem 1.5.

In the limit $l\to\infty$, we have $\mathcal{P}(l)\to 0$, so that $dP_{\text{ohmic}}/dA = \mathcal{P}(0)$, which states that all the power that enters at $z = 0$ will be dissipated into heat inside the semi-infinite medium. Using Eq. (2.6.17), we summarize this case:

$$\boxed{\frac{dP_{\text{ohmic}}}{dA} = \frac{1}{2}\,\text{Re}\big(\eta_c^{-1}\big)\,|\boldsymbol{E}_0|^2 = \frac{1}{2}\eta'\,|\boldsymbol{H}_0|^2} \qquad \text{(ohmic losses)} \tag{2.6.26}$$

This result will be used later on to calculate ohmic losses of waves incident on lossy dielectric or conductor surfaces, as well as conductor losses in waveguide and transmission line problems.

Example 2.6.1: The absorption coefficient α of water reaches a minimum over the visible spectrum—a fact undoubtedly responsible for why the visible spectrum is visible.

Recent measurements [113] of the absorption coefficient show that it starts at about 0.01 nepers/m at 380 nm (violet), decreases to a minimum value of 0.0044 nepers/m at 418 nm (blue), and then increases steadily reaching the value of 0.5 nepers/m at 600 nm (red). Determine the penetration depth δ in meters, for each of the three wavelengths.

Determine the depth in meters at which the light intensity has decreased to 1/10th its value at the surface of the water. Repeat, if the intensity is decreased to 1/100th its value.

Solution: The penetration depths $\delta = 1/\alpha$ are:

$$\delta = 100,\ 227.3,\ 2\ \text{m} \quad \text{for} \quad \alpha = 0.01,\ 0.0044,\ 0.5\ \text{nepers/m}$$

Using Eq. (2.6.21), we may solve for the depth $z = (A/8.9696)\,\delta$. Since a decrease of the light intensity (power) by a factor of 10 is equivalent to $A = 10$ dB, we find $z = (10/8.9696)\,\delta = 1.128\,\delta$, which gives: $z = 112.8,\ 256.3,\ 2.3$ m. A decrease by a factor of $100 = 10^{20/10}$ corresponds to $A = 20$ dB, effectively doubling the above depths. □

Example 2.6.2: A microwave oven operating at 2.45 GHz is used to defrost a frozen food having complex permittivity $\epsilon_c = (4 - j)\epsilon_0$ farad/m. Determine the strength of the electric field at a depth of 1 cm and express it in dB and as a percentage of its value at the surface. Repeat if $\epsilon_c = (45 - 15j)\epsilon_0$ farad/m.

Solution: The free-space wavenumber is $k_0 = \omega\sqrt{\mu_0\epsilon_0} = 2\pi f/c_0 = 2\pi(2.45{\times}10^9)/(3{\times}10^8) = 51.31$ rad/m. Using $k_c = \omega\sqrt{\mu_0\epsilon_c} = k_0\sqrt{\epsilon_c/\epsilon_0}$, we calculate the wavenumbers:

$$\begin{aligned} k_c &= \beta - j\alpha = 51.31\sqrt{4 - j} = 51.31(2.02 - 0.25j) = 103.41 - 12.73j\ \text{m}^{-1} \\ k_c &= \beta - j\alpha = 51.31\sqrt{45 - 15j} = 51.31(6.80 - 1.10j) = 348.84 - 56.61j\ \text{m}^{-1} \end{aligned}$$

The corresponding attenuation constants and penetration depths are:

$$\begin{aligned} \alpha &= 12.73\ \text{nepers/m}, & \delta &= 7.86\ \text{cm} \\ \alpha &= 56.61\ \text{nepers/m}, & \delta &= 1.77\ \text{cm} \end{aligned}$$

It follows that the attenuations at 1 cm will be in dB and in absolute units:

$$\begin{aligned} A &= 8.686\,z/\delta = 1.1\ \text{dB}, & 10^{-A/20} &= 0.88 \\ A &= 8.686\,z/\delta = 4.9\ \text{dB}, & 10^{-A/20} &= 0.57 \end{aligned}$$

Thus, the fields at a depth of 1 cm are 88% and 57% of their values at the surface. The complex permittivities of some foods may be found in [114]. □

A convenient way to characterize the degree of ohmic losses is by means of the *loss tangent*, originally defined in Eq. (1.9.35). Here, we set:

$$\tau = \tan\theta = \frac{\epsilon''}{\epsilon'} = \frac{\sigma + \omega\epsilon_d''}{\omega\epsilon_d'} \tag{2.6.27}$$

Then, $\epsilon_c = \epsilon' - j\epsilon'' = \epsilon'(1 - j\tau) = \epsilon_d'(1 - j\tau)$. Therefore, k_c, η_c may be written as:

$$k_c = \omega\sqrt{\mu\epsilon_d'}\,(1 - j\tau)^{1/2}, \quad \eta_c = \sqrt{\frac{\mu}{\epsilon_d'}}\,(1 - j\tau)^{-1/2} \tag{2.6.28}$$

The quantities $c_d = 1/\sqrt{\mu\epsilon_d'}$ and $\eta_d = \sqrt{\mu/\epsilon_d'}$ would be the speed of light and characteristic impedance of an equivalent lossless dielectric with permittivity ϵ_d'.

In terms of the loss tangent, we may characterize *weakly* lossy media versus *strongly* lossy ones by the conditions $\tau \ll 1$ versus $\tau \gg 1$, respectively. These conditions depend on the operating frequency ω:

$$\frac{\sigma + \omega\epsilon_d''}{\omega\epsilon_d'} \ll 1 \quad \text{versus} \quad \frac{\sigma + \omega\epsilon_d''}{\omega\epsilon_d'} \gg 1$$

The expressions (2.6.28) may be simplified considerably in these two limits. Using the small-x Taylor series expansion $(1 + x)^{1/2} \simeq 1 + x/2$, we find in the weakly lossy case $(1 - j\tau)^{1/2} \simeq 1 - j\tau/2$, and similarly, $(1 - j\tau)^{-1/2} \simeq 1 + j\tau/2$.

On the other hand, if $\tau \gg 1$, we may approximate $(1-j\tau)^{1/2} \simeq (-j\tau)^{1/2} = e^{-j\pi/4}\tau^{1/2}$, where we wrote $(-j)^{1/2} = (e^{-j\pi/2})^{1/2} = e^{-j\pi/4}$. Similarly, $(1 - j\tau)^{-1/2} \simeq e^{j\pi/4}\tau^{-1/2}$. Thus, we summarize the two limits:

$$(1 - j\tau)^{1/2} = \begin{cases} 1 - j\dfrac{\tau}{2}, & \text{if} \quad \tau \ll 1 \\ e^{-j\pi/4}\,\tau^{1/2} = (1 - j)\sqrt{\dfrac{\tau}{2}}, & \text{if} \quad \tau \gg 1 \end{cases} \tag{2.6.29}$$

$$(1 - j\tau)^{-1/2} = \begin{cases} 1 + j\dfrac{\tau}{2}, & \text{if} \quad \tau \ll 1 \\ e^{j\pi/4}\,\tau^{-1/2} = (1 + j)\sqrt{\dfrac{1}{2\tau}}, & \text{if} \quad \tau \gg 1 \end{cases} \tag{2.6.30}$$

2.7 Propagation in Weakly Lossy Dielectrics

In the weakly lossy case, the propagation parameters k_c, η_c become:

$$\begin{aligned} k_c &= \beta - j\alpha = \omega\sqrt{\mu\epsilon_d'}\left(1 - j\frac{\tau}{2}\right) = \omega\sqrt{\mu\epsilon_d'}\left(1 - j\frac{\sigma + \omega\epsilon_d''}{2\omega\epsilon_d'}\right) \\ \eta_c &= \eta' + j\eta'' = \sqrt{\frac{\mu}{\epsilon_d'}}\left(1 + j\frac{\tau}{2}\right) = \sqrt{\frac{\mu}{\epsilon_d'}}\left(1 + j\frac{\sigma + \omega\epsilon_d''}{2\omega\epsilon_d'}\right) \end{aligned} \tag{2.7.1}$$

Thus, the phase and attenuation constants are:

$$\beta = \omega\sqrt{\mu\epsilon_d'} = \frac{\omega}{c_d}, \quad \alpha = \frac{1}{2}\sqrt{\frac{\mu}{\epsilon_d'}}(\sigma + \omega\epsilon_d'') = \frac{1}{2}\eta_d(\sigma + \omega\epsilon_d'') \tag{2.7.2}$$

For a slightly conducting dielectric with $\epsilon_d'' = 0$ and a small conductivity σ, Eq. (2.7.2) implies that the attenuation coefficient α is frequency-independent in this limit.

Example 2.7.1: Seawater has $\sigma = 4$ Siemens/m and $\epsilon_d = 81\epsilon_0$ (so that $\epsilon_d' = 81\epsilon_0$, $\epsilon_d'' = 0$.) Then, $n_d = \sqrt{\epsilon_d/\epsilon_0} = 9$, and $c_d = c_0/n_d = 0.33\times 10^8$ m/sec and $\eta_d = \eta_0/n_d = 377/9 = 41.89\ \Omega$. The attenuation coefficient (2.7.2) will be:

$$\alpha = \frac{1}{2}\eta_d\sigma = \frac{1}{2}41.89\times 4 = 83.78 \text{ nepers/m} \quad\Rightarrow\quad \alpha_{\text{dB}} = 8.686\,\alpha = 728 \text{ dB/m}$$

The corresponding skin depth is $\delta = 1/\alpha = 1.19$ cm. This result assumes that $\sigma \ll \omega\epsilon_d$, which can be written in the form $\omega \gg \sigma/\epsilon_d$, or $f \gg f_0$, where $f_0 = \sigma/(2\pi\epsilon_d)$. Here, we have $f_0 = 888$ MHz. For frequencies $f \lesssim f_0$, we must use the exact equations (2.6.28). For example, we find:

$$\begin{array}{lll} f = 1\text{ kHz}, & \alpha_{\text{dB}} = 1.09\text{ dB/m}, & \delta = 7.96\text{ m} \\ f = 1\text{ MHz}, & \alpha_{\text{dB}} = 34.49\text{ dB/m}, & \delta = 25.18\text{ cm} \\ f = 1\text{ GHz}, & \alpha_{\text{dB}} = 672.69\text{ dB/m}, & \delta = 1.29\text{ cm} \end{array}$$

Such extremely large attenuations explain why communication with submarines is impossible at high RF frequencies. □

2.8 *Propagation in Good Conductors*

A conductor is characterized by a large value of its conductivity σ, while its dielectric constant may be assumed to be real-valued $\epsilon_d = \epsilon$ (typically equal to ϵ_0.) Thus, its complex permittivity and loss tangent will be:

$$\epsilon_c = \epsilon - j\frac{\sigma}{\omega} = \epsilon\left(1 - j\frac{\sigma}{\omega\epsilon}\right), \qquad \tau = \frac{\sigma}{\omega\epsilon} \tag{2.8.1}$$

A good conductor corresponds to the limit $\tau \gg 1$, or, $\sigma \gg \omega\epsilon$. Using the approximations of Eqs. (2.6.29) and (2.6.30), we find for the propagation parameters k_c, η_c:

$$\begin{aligned} k_c &= \beta - j\alpha = \omega\sqrt{\mu\epsilon}\sqrt{\frac{\tau}{2}}\,(1-j) = \sqrt{\frac{\omega\mu\sigma}{2}}\,(1-j) \\ \eta_c &= \eta' + j\eta'' = \sqrt{\frac{\mu}{\epsilon}}\sqrt{\frac{1}{2\tau}}\,(1+j) = \sqrt{\frac{\omega\mu}{2\sigma}}\,(1+j) \end{aligned} \tag{2.8.2}$$

Thus, the parameters β, α, δ are:

$$\boxed{\beta = \alpha = \sqrt{\frac{\omega\mu\sigma}{2}} = \sqrt{\pi f\mu\sigma}}\,, \qquad \boxed{\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu\sigma}} = \frac{1}{\sqrt{\pi f\mu\sigma}}} \tag{2.8.3}$$

where we replaced $\omega = 2\pi f$. The complex characteristic impedance η_c can be written in the form $\eta_c = R_s(1+j)$, where R_s is called the *surface resistance* and is given by the equivalent forms (where $\eta = \sqrt{\mu/\epsilon}$):

$$\boxed{R_s = \eta\sqrt{\frac{\omega\epsilon}{2\sigma}} = \sqrt{\frac{\omega\mu}{2\sigma}} = \frac{\alpha}{\sigma} = \frac{1}{\sigma\delta}} \tag{2.8.4}$$

Example 2.8.1: For copper we have $\sigma = 5.8\times 10^7$ Siemens/m. The skin depth at frequency f is:

$$\delta = \frac{1}{\sqrt{\pi f\mu\sigma}} = \frac{1}{\sqrt{\pi\cdot 4\pi\cdot 10^{-7}\cdot 5.8\cdot 10^7}}\,f^{-1/2} = 0.0661\,f^{-1/2} \qquad (f \text{ in Hz})$$

We find at frequencies of 1 kHz, 1 MHz, and 1 GHz:

$$\begin{array}{ll} f = 1\text{ kHz}, & \delta = 2.09\text{ mm} \\ f = 1\text{ MHz}, & \delta = 0.07\text{ mm} \\ f = 1\text{ GHz}, & \delta = 2.09\ \mu\text{m} \end{array}$$

Thus, the skin depth is extremely small for good conductors at RF. □

Because δ is so small, the fields will attenuate rapidly within the conductor, depending on distance like $e^{-\gamma z} = e^{-\alpha z}e^{-j\beta z} = e^{-z/\delta}e^{-j\beta z}$. The factor $e^{-z/\delta}$ effectively confines the fields to within a distance δ from the surface of the conductor.

This allows us to define equivalent "surface" quantities, such as surface current and surface impedance. With reference to Fig. 2.6.1, we define the surface current density by integrating the density $\mathbf{J}(z) = \sigma\mathbf{E}(z) = \sigma\mathbf{E}_0 e^{-\gamma z}$ over the top-side of the volume $l\,dA$, and taking the limit $l \to \infty$:

$$\mathbf{J}_s = \int_0^\infty \mathbf{J}(z)\,dz = \int_0^\infty \sigma\mathbf{E}_0 e^{-\gamma z}dz = \frac{\sigma}{\gamma}\mathbf{E}_0\,, \quad \text{or,}$$

$$\mathbf{J}_s = \frac{1}{Z_s}\,\mathbf{E}_0 \tag{2.8.5}$$

where we defined the *surface impedance* $Z_s = \gamma/\sigma$. In the good-conductor limit, Z_s is equal to η_c. Indeed, it follows from Eqs. (2.8.3) and (2.8.4) that:

$$Z_s = \frac{\gamma}{\sigma} = \frac{\alpha + j\beta}{\sigma} = \frac{\alpha}{\sigma}(1+j) = R_s(1+j) = \eta_c$$

Because $\mathbf{H}_0 \times \hat{\mathbf{z}} = \mathbf{E}_0/\eta_c$, it follows that the surface current will be related to the magnetic field intensity at the surface of the conductor by:

$$\mathbf{J}_s = \mathbf{H}_0 \times \hat{\mathbf{z}} = \hat{\mathbf{n}} \times \mathbf{H}_0 \tag{2.8.6}$$

where $\hat{\mathbf{n}} = -\hat{\mathbf{z}}$ is the outward normal to the conductor. The meaning of $\mathbf{J}_s$ is that it represents the current flowing in the direction of $\mathbf{E}_0$ per unit length measured along the perpendicular direction to $\mathbf{E}_0$, that is, the $\mathbf{H}_0$-direction. It has units of A/m.

The total amount of ohmic losses per unit surface area of the conductor may be calculated from Eq. (2.6.26), which reads in this case:

$$\boxed{\frac{dP_{\text{ohmic}}}{dA} = \frac{1}{2}R_s|\mathbf{H}_0|^2 = \frac{1}{2}R_s|\mathbf{J}_s|^2} \quad \text{(ohmic loss per unit conductor area)} \tag{2.8.7}$$

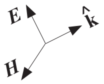
2.9 Propagation in Oblique Directions

So far we considered waves propagating towards the z-direction. For single-frequency uniform plane waves propagating in some arbitrary direction in a lossless medium, the propagation factor is obtained by the substitution:

$$e^{-jkz} \to e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

where $\boldsymbol{k} = k\hat{\mathbf{k}}$, with $k = \omega\sqrt{\mu\epsilon} = \omega/c$ and $\hat{\mathbf{k}}$ is a *unit vector* in the direction of propagation. The fields take the form:

$$\begin{aligned} \boldsymbol{E}(\mathbf{r},t) &= \boldsymbol{E}_0 e^{j\omega t - j\boldsymbol{k}\cdot\mathbf{r}} \\ \boldsymbol{H}(\mathbf{r},t) &= \boldsymbol{H}_0 e^{j\omega t - j\boldsymbol{k}\cdot\mathbf{r}} \end{aligned} \tag{2.9.1}$$



where $\boldsymbol{E}_0$, $\boldsymbol{H}_0$ are constant vectors transverse to $\hat{\mathbf{k}}$, that is, $\hat{\mathbf{k}}\cdot\boldsymbol{E}_0 = \hat{\mathbf{k}}\cdot\boldsymbol{H}_0 = 0$, such that:

$$\boldsymbol{H}_0 = \frac{1}{\omega\mu}\,\boldsymbol{k}\times\boldsymbol{E}_0 = \frac{1}{\eta}\,\hat{\mathbf{k}}\times\boldsymbol{E}_0 \tag{2.9.2}$$

where $\eta = \sqrt{\mu/\epsilon}$. Thus, $\{\boldsymbol{E},\boldsymbol{H},\hat{\mathbf{k}}\}$ form a right-handed orthogonal system.

The solutions (2.9.1) can be derived from Maxwell's equations in a straightforward fashion. When the gradient operator acts on the above fields, it can be simplified into $\nabla \to -j\boldsymbol{k}$. This follows from:

$$\nabla\left(e^{-j\boldsymbol{k}\cdot\mathbf{r}}\right) = -j\boldsymbol{k}\left(e^{-j\boldsymbol{k}\cdot\mathbf{r}}\right)$$

After canceling the common factor $e^{j\omega t - j\boldsymbol{k}\cdot\mathbf{r}}$, Maxwell's equations (2.1.1) take the form:

$$\begin{array}{lcl} -j\boldsymbol{k}\times\boldsymbol{E}_0 = -j\omega\mu\boldsymbol{H}_0 & & \boldsymbol{k}\times\boldsymbol{E}_0 = \omega\mu\boldsymbol{H}_0 \\ -j\boldsymbol{k}\times\boldsymbol{H}_0 = j\omega\epsilon\boldsymbol{E}_0 & \Rightarrow & \boldsymbol{k}\times\boldsymbol{H}_0 = -\omega\epsilon\boldsymbol{E}_0 \\ \boldsymbol{k}\cdot\boldsymbol{E}_0 = 0 & & \boldsymbol{k}\cdot\boldsymbol{E}_0 = 0 \\ \boldsymbol{k}\cdot\boldsymbol{H}_0 = 0 & & \boldsymbol{k}\cdot\boldsymbol{H}_0 = 0 \end{array} \tag{2.9.3}$$

The last two imply that $\boldsymbol{E}_0, \boldsymbol{H}_0$ are transverse to $\boldsymbol{k}$. The other two can be decoupled by taking the cross product of the first equation with $\boldsymbol{k}$ and using the second equation:

$$\boldsymbol{k}\times(\boldsymbol{k}\times\boldsymbol{E}_0) = \omega\mu\,\boldsymbol{k}\times\boldsymbol{H}_0 = -\omega^2\mu\epsilon\,\boldsymbol{E}_0 \tag{2.9.4}$$

The left-hand side can be simplified using the BAC-CAB rule and $\boldsymbol{k}\cdot\boldsymbol{E}_0 = 0$, that is, $\boldsymbol{k}\times(\boldsymbol{k}\times\boldsymbol{E}_0) = \boldsymbol{k}(\boldsymbol{k}\cdot\boldsymbol{E}_0) - \boldsymbol{E}_0(\boldsymbol{k}\cdot\boldsymbol{k}) = -(\boldsymbol{k}\cdot\boldsymbol{k})\boldsymbol{E}_0$. Thus, Eq. (2.9.4) becomes:

$$-(\boldsymbol{k}\cdot\boldsymbol{k})\,\boldsymbol{E}_0 = -\omega^2\mu\epsilon\,\boldsymbol{E}_0$$

Thus, we obtain the consistency condition:

$$\boxed{\boldsymbol{k}\cdot\boldsymbol{k} = \omega^2\mu\epsilon} \tag{2.9.5}$$

Defining $k = \sqrt{\boldsymbol{k}\cdot\boldsymbol{k}} = |\boldsymbol{k}|$, we have $k = \omega\sqrt{\mu\epsilon}$. Using the relationship $\omega\mu = k\eta$ and defining the unit vector $\hat{\mathbf{k}} = \boldsymbol{k}/|\boldsymbol{k}| = \boldsymbol{k}/k$, the magnetic field is obtained from:

$$\boldsymbol{H}_0 = \frac{\boldsymbol{k}\times\boldsymbol{E}_0}{\omega\mu} = \frac{\boldsymbol{k}\times\boldsymbol{E}_0}{k\eta} = \frac{1}{\eta}\,\hat{\mathbf{k}}\times\boldsymbol{E}_0$$

The constant-phase (and constant-amplitude) wavefronts are the planes $\boldsymbol{k}\cdot\mathbf{r}$ = constant, or, $\hat{\mathbf{k}}\cdot\mathbf{r}$ = constant. They are the planes perpendicular to the propagation direction $\hat{\mathbf{k}}$.

As an example, consider a rotated coordinate system $\{x', y', z'\}$ in which the $z'x'$ axes are rotated by angle θ relative to the original zx axes, as shown in Fig. 2.9.1. Thus, the new coordinates and corresponding unit vectors will be:

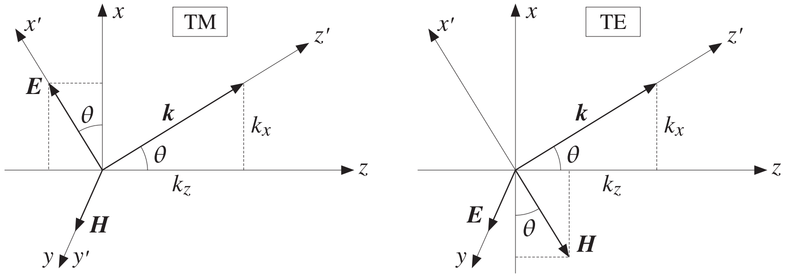


Fig. 2.9.1 TM and TE waves.

$$\begin{array}{ll} z' = z\cos\theta + x\sin\theta, & \hat{\mathbf{z}}' = \hat{\mathbf{z}}\cos\theta + \hat{\mathbf{x}}\sin\theta \\ x' = x\cos\theta - z\sin\theta, & \hat{\mathbf{x}}' = \hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta \\ y' = y, & \hat{\mathbf{y}}' = \hat{\mathbf{y}} \end{array} \tag{2.9.6}$$

We choose the propagation direction to be the new z-axis, that is, $\hat{\mathbf{k}} = \hat{\mathbf{z}}'$, so that the wave vector $\boldsymbol{k} = k\,\hat{\mathbf{k}} = k\,\hat{\mathbf{z}}'$ will have components $k_z = k\cos\theta$ and $k_x = k\sin\theta$:

$$\boldsymbol{k} = k\,\hat{\mathbf{k}} = k\,(\hat{\mathbf{z}}\cos\theta + \hat{\mathbf{x}}\sin\theta) = \hat{\mathbf{z}}\,k_z + \hat{\mathbf{x}}\,k_x$$

The propagation phase factor becomes:

$$e^{-j\boldsymbol{k}\cdot\mathbf{r}} = e^{-j(k_z z + k_x x)} = e^{-jk(z\cos\theta + x\sin\theta)} = e^{-jkz'}$$

Because $\{\boldsymbol{E}_0, \boldsymbol{H}_0, \boldsymbol{k}\}$ form a right-handed vector system, the electric field may have components along the new transverse (with respect to z') axes, that is, along x' and y. Thus, we may resolve $\boldsymbol{E}_0$ into the orthogonal directions:

$$\boldsymbol{E}_0 = \hat{\mathbf{x}}'A + \hat{\mathbf{y}}\,B = (\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta)A + \hat{\mathbf{y}}\,B \tag{2.9.7}$$

The corresponding magnetic field will be $\boldsymbol{H}_0 = \hat{\mathbf{k}}\times\boldsymbol{E}_0/\eta = \hat{\mathbf{z}}'\times(\hat{\mathbf{x}}'A + \hat{\mathbf{y}}\,B)/\eta$. Using the relationships $\hat{\mathbf{z}}'\times\hat{\mathbf{x}}' = \hat{\mathbf{y}}$ and $\hat{\mathbf{z}}'\times\hat{\mathbf{y}} = -\hat{\mathbf{x}}'$, we find:

$$\boldsymbol{H}_0 = \frac{1}{\eta}\left[\hat{\mathbf{y}}\,A - \hat{\mathbf{x}}'B\right] = \frac{1}{\eta}\left[\hat{\mathbf{y}}\,A - (\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta)B\right] \tag{2.9.8}$$

The complete expressions for the fields are then:

$$\begin{aligned}\boldsymbol{E}(\mathbf{r},t) &= \big[(\hat{\mathbf{x}}\cos\theta-\hat{\mathbf{z}}\sin\theta)A+\hat{\mathbf{y}}\,B\big]e^{j\omega t-jk(z\cos\theta+x\sin\theta)}\\ \boldsymbol{H}(\mathbf{r},t) &= \frac{1}{\eta}\big[\hat{\mathbf{y}}\,A-(\hat{\mathbf{x}}\cos\theta-\hat{\mathbf{z}}\sin\theta)B\big]e^{j\omega t-jk(z\cos\theta+x\sin\theta)}\end{aligned}\tag{2.9.9}$$

Written with respect to the rotated coordinate system $\{x',y',z'\}$, the solutions become identical to those of Sec. 2.2:

$$\begin{aligned}\boldsymbol{E}(\mathbf{r},t) &= \big[\hat{\mathbf{x}}'A+\hat{\mathbf{y}}'B\big]e^{j\omega t-jkz'}\\ \boldsymbol{H}(\mathbf{r},t) &= \frac{1}{\eta}\big[\hat{\mathbf{y}}'A-\hat{\mathbf{x}}'B\big]e^{j\omega t-jkz'}\end{aligned}\tag{2.9.10}$$

They are uniform in the sense that they do not depend on the new transverse coordinates x',y'. The constant-phase planes are $z'=\hat{\mathbf{z}}'\cdot\mathbf{r}=z\cos\theta+x\sin\theta=$ const.

The polarization properties of the wave depend on the relative phases and amplitudes of the complex constants A,B, with the polarization ellipse lying on the $x'y'$ plane.

The A- and B-components of $\boldsymbol{E}_0$ are referred to as *transverse magnetic* (TM) and *transverse electric* (TE), respectively, where "transverse" is meant here with respect to the z-axis. The TE case has an *electric* field transverse to z; the TM case has a *magnetic* field transverse to z. Fig. 2.9.1 depicts these two cases separately.

This nomenclature arises in the context of plane waves incident obliquely on interfaces, where the xz plane is the plane of incidence and the interface is the xy plane. The TE and TM cases are also referred to as having "perpendicular" and "parallel" polarization vectors with respect to the plane of incidence, that is, the E-field is perpendicular or parallel to the xz plane.

We may define the concept of *transverse impedance* as the ratio of the transverse (with respect to z) components of the electric and magnetic fields. In particular, by analogy with the definitions of Sec. 2.4, we have:

$$\begin{aligned}\eta_{TM} &= \frac{E_x}{H_y}=\frac{A\cos\theta}{\dfrac{1}{\eta}A}=\eta\cos\theta\\ \eta_{TE} &= -\frac{E_y}{H_x}=\frac{B}{\dfrac{1}{\eta}B\cos\theta}=\frac{\eta}{\cos\theta}\end{aligned}\tag{2.9.11}$$

Such transverse impedances play an important role in describing the transfer matrices of dielectric slabs at oblique incidence. We discuss them further in Chap. 7.

2.10 Complex or Inhomogeneous Waves

The steps leading to the wave solution (2.9.1) do not preclude a *complex-valued* wavevector $\boldsymbol{k}$. For example, if the medium is lossy, we must replace $\{\eta,k\}$ by $\{\eta_c,k_c\}$, where $k_c=\beta-j\alpha$, resulting from a complex effective permittivity ϵ_c. If the propagation direction is defined by the unit vector $\hat{\mathbf{k}}$, chosen to be a *rotated* version of $\hat{\mathbf{z}}$, then the wavevector will be defined by $\boldsymbol{k}=k_c\hat{\mathbf{k}}=(\beta-j\alpha)\hat{\mathbf{k}}$. Because $k_c=\omega\sqrt{\mu\epsilon_c}$ and $\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}=1$, the vector $\boldsymbol{k}$ satisfies the consistency condition (2.9.5):

$$\boldsymbol{k}\cdot\boldsymbol{k}=k_c^2=\omega^2\mu\epsilon_c\tag{2.10.1}$$

The propagation factor will be:

$$e^{-j\boldsymbol{k}\cdot\mathbf{r}}=e^{-jk_c\hat{\mathbf{k}}\cdot\mathbf{r}}=e^{-(\alpha+j\beta)\,\hat{\mathbf{k}}\cdot\mathbf{r}}=e^{-\alpha\hat{\mathbf{k}}\cdot\mathbf{r}}e^{-j\beta\hat{\mathbf{k}}\cdot\mathbf{r}}$$

The wave is still a uniform plane wave in the sense that the constant-amplitude planes, $\alpha\hat{\mathbf{k}}\cdot\mathbf{r}=$ const., and the constant-phase planes, $\beta\hat{\mathbf{k}}\cdot\mathbf{r}=$ const., coincide with each other—being the planes perpendicular to the propagation direction. For example, the rotated solution (2.9.10) becomes in the lossy case:

$$\begin{aligned}\boldsymbol{E}(\mathbf{r},t) &= \big[\hat{\mathbf{x}}'A+\hat{\mathbf{y}}'B\big]e^{j\omega t-jk_cz'}=\big[\hat{\mathbf{x}}'A+\hat{\mathbf{y}}'B\big]e^{j\omega t-(\alpha+j\beta)z'}\\ \boldsymbol{H}(\mathbf{r},t) &= \frac{1}{\eta_c}\big[\hat{\mathbf{y}}'A-\hat{\mathbf{x}}'B\big]e^{j\omega t-jk_cz'}=\frac{1}{\eta_c}\big[\hat{\mathbf{y}}'A-\hat{\mathbf{x}}'B\big]e^{j\omega t-(\alpha+j\beta)z'}\end{aligned}\tag{2.10.2}$$

In this solution, the real and imaginary parts of the wavevector $\boldsymbol{k}=\boldsymbol{\beta}-j\boldsymbol{\alpha}$ are collinear, that is, $\boldsymbol{\beta}=\beta\hat{\mathbf{k}}$ and $\boldsymbol{\alpha}=\alpha\hat{\mathbf{k}}$.

More generally, there exist solutions having a complex wavevector $\boldsymbol{k}=\boldsymbol{\beta}-j\boldsymbol{\alpha}$ such that $\boldsymbol{\beta},\boldsymbol{\alpha}$ are not collinear. The propagation factor becomes now:

$$e^{-j\boldsymbol{k}\cdot\mathbf{r}}=e^{-(\boldsymbol{\alpha}+j\boldsymbol{\beta})\cdot\mathbf{r}}=e^{-\boldsymbol{\alpha}\cdot\mathbf{r}}e^{-j\boldsymbol{\beta}\cdot\mathbf{r}}\tag{2.10.3}$$

If $\boldsymbol{\alpha},\boldsymbol{\beta}$ are not collinear, such a wave will not be a uniform plane wave because the constant-amplitude planes, $\boldsymbol{\alpha}\cdot\mathbf{r}=$ const., and the constant-phase planes, $\boldsymbol{\beta}\cdot\mathbf{r}=$ const., will be different. The consistency condition $\boldsymbol{k}\cdot\boldsymbol{k}=k_c^2=(\beta-j\alpha)^2$ splits into the following two conditions obtained by equating real and imaginary parts:

$$(\boldsymbol{\beta}-j\boldsymbol{\alpha})\cdot(\boldsymbol{\beta}-j\boldsymbol{\alpha})=(\beta-j\alpha)^2\quad\Leftrightarrow\quad\boxed{\begin{aligned}\boldsymbol{\beta}\cdot\boldsymbol{\beta}-\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}&=\beta^2-\alpha^2\\ \boldsymbol{\beta}\cdot\boldsymbol{\alpha}&=\alpha\beta\end{aligned}}\tag{2.10.4}$$

With $\boldsymbol{E}_0$ chosen to satisfy $\boldsymbol{k}\cdot\boldsymbol{E}_0=(\boldsymbol{\beta}-j\boldsymbol{\alpha})\cdot\boldsymbol{E}_0=0$, the magnetic field is computed from Eq. (2.9.2), $\boldsymbol{H}_0=\boldsymbol{k}\times\boldsymbol{E}_0/\omega\mu=(\boldsymbol{\beta}-j\boldsymbol{\alpha})\times\boldsymbol{E}_0/\omega\mu$.

Let us look at an explicit construction. We choose $\boldsymbol{\beta},\boldsymbol{\alpha}$ to lie on the xz plane of Fig. 2.9.1, and resolve them as $\boldsymbol{\beta}=\hat{\mathbf{z}}\,\beta_z+\hat{\mathbf{x}}\,\beta_x$ and $\boldsymbol{\alpha}=\hat{\mathbf{z}}\,\alpha_z+\hat{\mathbf{x}}\,\alpha_x$. Thus,

$$\boldsymbol{k}=\boldsymbol{\beta}-j\boldsymbol{\alpha}=\hat{\mathbf{z}}\,(\beta_z-j\alpha_z)+\hat{\mathbf{x}}\,(\beta_x-j\alpha_x)=\hat{\mathbf{z}}\,k_z+\hat{\mathbf{x}}\,k_x$$

Then, the propagation factor (2.10.3) and conditions (2.10.4) read explicitly:

$$\begin{aligned}&e^{-j\boldsymbol{k}\cdot\mathbf{r}}=e^{-(\alpha_zz+\alpha_xx)}e^{-j(\beta_zz+\beta_xx)}\\ &\beta_z^2+\beta_x^2-\alpha_z^2-\alpha_x^2=\beta^2-\alpha^2\\ &\beta_z\alpha_z+\beta_x\alpha_x=\beta\alpha\end{aligned}\tag{2.10.5}$$

Because $\mathbf{k}$ is orthogonal to both $\hat{\mathbf{y}}$ and $\hat{\mathbf{y}}\times\mathbf{k}$, we construct the electric field $\mathbf{E}_0$ as the following linear combination of TM and TE terms:

$$\mathbf{E}_0 = (\hat{\mathbf{y}}\times\hat{\mathbf{k}})A + \hat{\mathbf{y}}B\,, \qquad \text{where} \quad \hat{\mathbf{k}} = \frac{\mathbf{k}}{k_c} = \frac{\boldsymbol{\beta}-j\boldsymbol{\alpha}}{\beta - j\alpha} \tag{2.10.6}$$

This satisfies $\mathbf{k}\cdot\mathbf{E}_0 = 0$. Then, the magnetic field becomes:

$$\mathbf{H}_0 = \frac{\mathbf{k}\times\mathbf{E}_0}{\omega\mu} = \frac{1}{\eta_c}\big[\hat{\mathbf{y}}A - (\hat{\mathbf{y}}\times\hat{\mathbf{k}})B\big] \tag{2.10.7}$$

The vector $\hat{\mathbf{k}}$ is complex-valued and satisfies $\hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1$. These expressions reduce to Eq. (2.10.2), if $\hat{\mathbf{k}} = \hat{\mathbf{z}}'$.

Waves with a complex $\mathbf{k}$ are known as *complex waves*, or *inhomogeneous waves*. In applications, they always appear in connection with some interface between two media. The interface serves either as a reflecting/transmitting surface, or as a guiding surface.

For example, when plane waves are incident obliquely from a lossless dielectric onto a planar interface with a lossy medium, the waves transmitted into the lossy medium are of such complex type. Taking the interface to be the xy-plane and the lossy medium to be the region $z \geq 0$, it turns out that the transmitted waves are characterized by attenuation only in the z-direction. Therefore, Eqs. (2.10.5) apply with $\alpha_z > 0$ and $\alpha_x = 0$. The parameter β_x is fixed by Snel's law, so that Eqs. (2.10.5) provide a system of two equations in the two unknowns β_z and α_z. We discuss this further in Chap. 7.

Wave solutions with complex $\mathbf{k} = \boldsymbol{\beta} - j\boldsymbol{\alpha}$ are possible even when the propagation medium is lossless so that $\epsilon_c = \epsilon$ is real, and $\beta = \omega\sqrt{\mu\epsilon}$ and $\alpha = 0$. Then, Eqs. (2.10.4) become $\boldsymbol{\beta}\cdot\boldsymbol{\beta} - \boldsymbol{\alpha}\cdot\boldsymbol{\alpha} = \beta^2$ and $\boldsymbol{\beta}\cdot\boldsymbol{\alpha} = 0$. Thus, the constant-amplitude and constant-phase planes are *orthogonal* to each other.

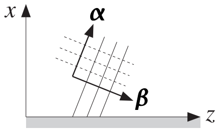
Examples of such waves are the evanescent waves in total internal reflection, various guided-wave problems, such as surface waves, leaky waves, and traveling-wave antennas. The most famous of these is the Zenneck wave, which is a surface wave propagating along a lossy ground, decaying exponentially with distance above and along the ground.

Another example of current interest is surface plasmons [553–591], which are surface waves propagating along the interface between a metal, such as silver, and a dielectric, such as air, with the fields decaying exponentially perpendicularly to the interface both in the air and the metal. We discuss them further in Sections 7.11 and 8.5.

For a classification of various types of complex waves and a review of several applications, including the Zenneck wave, see Refs. [869–876]. We will encounter some of these in Section 7.7.

The table below illustrates the vectorial directions and relative signs of some possible types, assuming that $\boldsymbol{\alpha}, \boldsymbol{\beta}$ lie on the xz plane with the yz plane being the interface plane.

$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	α_z	α_x	β_z	β_x	complex wave type
0	↘	0	0	+	−	oblique incidence
↑	→	0	+	+	0	evanescent surface wave
↗	↘	+	+	+	−	Zenneck surface wave
↖	↗	−	+	+	+	leaky wave



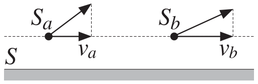
2.11 Doppler Effect

The Doppler effect is the frequency shift perceived by an observer whenever the source of the waves and the observer are in relative motion.

Besides the familiar Doppler effect for sound waves, such as the increase in pitch of the sound of an approaching car, ambulance, or train, the Doppler effect has several other applications, such as Doppler radar for aircraft tracking, the weather, ground imaging, and police radar; several medical ultrasound applications, such as monitoring blood flow or imaging internal organs and fetuses; and astrophysical applications, such as measuring the red shift of light emitted by receding galaxies.

In the classical treatment of the Doppler effect, one assumes that the waves propagate in some medium (e.g., sound waves in air). If c is the wave propagation speed in the medium, the classical expression for the Doppler effect is given by:

$$\boxed{f_b = f_a\frac{c - v_b}{c - v_a}} \tag{2.11.1}$$



where f_a and f_b are the frequencies measured in the rest frames of the source S_a and observer S_b, and v_a and v_b are the velocities of S_a and S_b with respect to the propagation medium and projected along their line of sight.

The algebraic sign of v_a is positive if S_a is moving toward S_b from the left, and the sign of v_b is positive if S_b is moving away from S_a. Thus, there is a frequency increase whenever the source and the observer are approaching each other ($v_a > 0$ or $v_b < 0$), and a frequency decrease if they are receding from each other ($v_a < 0$ or $v_b > 0$).

Eq. (2.11.1) can be derived by considering the two cases of a moving source and a stationary observer, or a stationary source and a moving observer, as shown in Fig. 2.11.1.

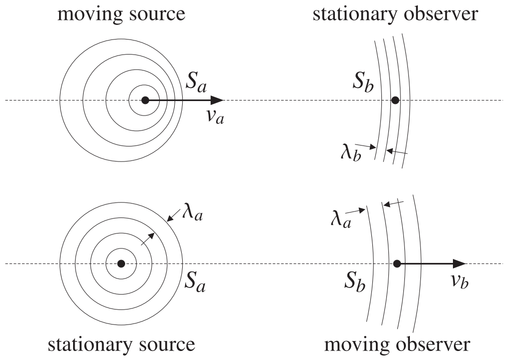


Fig. 2.11.1 Classical Doppler effect.

In the first case, the spacing of the successive crests of the wave (the wavelength) is decreased in front of the source because during the time interval between crests, that is, during one period $T_a = 1/f_a$, the source has moved by a distance v_aT_a bringing two successive crests closer together by that amount. Thus, the wavelength perceived by the observer will be $\lambda_b = \lambda_a - v_a T_a = (c - v_a)/f_a$, which gives:

$$\boxed{f_b = \frac{c}{\lambda_b} = f_a\frac{c}{c - v_a}}\qquad\text{(moving source)}\tag{2.11.2}$$

In the second case, because the source is stationary, the wavelength λ_a will not change, but now the effective speed of the wave in the rest frame of the observer is $(c - v_b)$. Therefore, the frequency perceived by the observer will be:

$$\boxed{f_b = \frac{c - v_b}{\lambda_a} = f_a\frac{c - v_b}{c}}\qquad\text{(moving observer)}\tag{2.11.3}$$

The combination of these two cases leads to Eq. (2.11.1). We have assumed in Eqs. (2.11.1)–(2.11.3) that v_a, v_b are less than c so that supersonic effects are not considered. A counter-intuitive aspect of the classical Doppler formula (2.11.1) is that it does not depend on the relative velocity $(v_b - v_a)$ of the observer and source. Therefore, it makes a difference whether the source or the observer is moving. Indeed, when the observer is moving with $v_b = v$ away from a stationary source, or when the source is moving with $v_a = -v$ away from a stationary observer, then Eq. (2.11.1) gives:

$$f_b = f_a(1 - v/c)\,,\qquad f_b = \frac{f_a}{1 + v/c}\tag{2.11.4}$$

These two expressions are equivalent to first-order in v/c. This follows from the Taylor series approximation $(1+x)^{-1} \simeq 1-x$, which is valid for $|x| \ll 1$. More generally, to first order in v_a/c and v_b/c, Eq. (2.11.1) does depend only on the relative velocity. In this case the Doppler shift $\Delta f = f_b - f_a$ is given approximately by:

$$\frac{\Delta f}{f_a} = \frac{v_a - v_b}{c}\tag{2.11.5}$$

For Doppler radar this doubles to $\Delta f/f_a = 2(v_a - v_b)/c$ because the wave suffers two Doppler shifts, one for the transmitted and one for the reflected wave. This is further discussed in Sec. 5.8.

For electromagnetic waves,[†] the correct Doppler formula depends only on the relative velocity between observer and source and is given by the relativistic generalization of Eq. (2.11.1):

$$\boxed{f_b = f_a\sqrt{\frac{c - v_b}{c + v_b}\cdot\frac{c + v_a}{c - v_a}} = f_a\sqrt{\frac{c - v}{c + v}}}\qquad\text{(relativistic Doppler effect)}\tag{2.11.6}$$

where v is the velocity of the observer *relative* to the source, which according to the Einstein addition theorem for velocities is given through the equivalent expressions:

$$v = \frac{v_b - v_a}{1 - v_b v_a/c^2}\quad\Leftrightarrow\quad v_b = \frac{v_a + v}{1 + v_a v/c^2}\quad\Leftrightarrow\quad \frac{c - v}{c + v} = \frac{c - v_b}{c + v_b}\cdot\frac{c + v_a}{c - v_a}\tag{2.11.7}$$

[†] The question of the existence of a medium (the ether) required for the propagation of electromagnetic waves precipitated the development of the special relativity theory.

Using the first-order Taylor series expansion $(1 + x)^{\pm 1/2} = 1 \pm x/2$, one can show that Eq. (2.11.6) can be written approximately as Eq. (2.11.5).

Next, we present a more precise discussion of the Doppler effect based on Lorentz transformations. Our discussion follows that of Einstein's 1905 paper on special relativity [435]. Fig. 2.11.2 shows a uniform plane wave propagating in vacuum as viewed from the vantage point of two coordinate frames: a fixed frame S and a frame S' moving towards the z-direction with velocity v. We assume that the wavevector $\mathbf{k}$ in S lies in the xz-plane and forms an angle θ with the z-axis as shown.

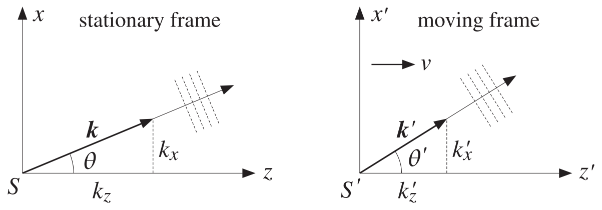


Fig. 2.11.2 Plane wave viewed from stationary and moving frames.

As discussed in Appendix H, the transformation of the frequency-wavenumber four-vector $(\omega/c, \mathbf{k})$ between the frames S and S' is given by the Lorentz transformation of Eq. (H.14). Because $k_y = 0$ and the transverse components of $\mathbf{k}$ do not change, we will have $k'_y = k_y = 0$, that is, the wavevector $\mathbf{k}'$ will still lie in the xz-plane of the S' frame. The frequency and the other components of $\mathbf{k}$ transform as follows:

$$\boxed{\begin{aligned} \omega' &= \gamma(\omega - \beta c k_z)\\ k'_z &= \gamma\left(k_z - \frac{\beta}{c}\omega\right)\\ k'_x &= k_x \end{aligned}}\qquad \beta = \frac{v}{c}\,,\quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}\tag{2.11.8}$$

Setting $k_z = k\cos\theta$, $k_x = k\sin\theta$, with $k = \omega/c$, and similarly in the S' frame, $k'_z = k'\cos\theta'$, $k'_x = k'\sin\theta'$, with $k' = \omega'/c$, Eqs. (2.11.8) may be rewritten in the form:

$$\begin{aligned} \omega' &= \omega\gamma(1 - \beta\cos\theta)\\ \omega'\cos\theta' &= \omega\gamma(\cos\theta - \beta)\\ \omega'\sin\theta' &= \omega\sin\theta \end{aligned}\tag{2.11.9}$$

The first equation is the relativistic Doppler formula, relating the frequency of the wave as it is measured by an observer in the moving frame S' to the frequency of a source in the fixed frame S:

$$\boxed{f' = f\gamma(1 - \beta\cos\theta) = f\,\frac{1 - \beta\cos\theta}{\sqrt{1 - \beta^2}}}\tag{2.11.10}$$

The last two equations in (2.11.9) relate the apparent propagation angles θ, θ' in the two frames. Eliminating ω, ω', we obtain the following equivalent expressions:

$$\cos\theta' = \frac{\cos\theta - \beta}{1 - \beta\cos\theta}\;\Leftrightarrow\;\sin\theta' = \frac{\sin\theta}{\gamma(1 - \beta\cos\theta)}\;\Leftrightarrow\;\frac{\tan(\theta'/2)}{\tan(\theta/2)} = \sqrt{\frac{1 + \beta}{1 - \beta}}\tag{2.11.11}$$

where to obtain the last one we used the identity $\tan(\phi/2)=\sin\phi/(1+\cos\phi)$. The difference in the propagation angles θ,θ' is referred to as the *aberration* of light due to motion. Using Eqs. (2.11.11), the Doppler equation (2.11.10) may be written in the alternative forms:

$$f' = f\gamma(1-\beta\cos\theta)=\frac{f}{\gamma(1+\beta\cos\theta')}=f\sqrt{\frac{1-\beta\cos\theta}{1+\beta\cos\theta'}} \tag{2.11.12}$$

If the wave is propagating in the z-direction ($\theta = 0^{\text{o}}$), Eq. (2.11.10) gives:

$$f' = f\sqrt{\frac{1-\beta}{1+\beta}} \tag{2.11.13}$$

and, if it is propagating in the x-direction ($\theta = 90^{\text{o}}$), we obtain the so-called *transverse* Doppler effect: $f' = f\gamma$. The relativistic Doppler effect, including the transverse one, has been observed experimentally.

To derive Eq. (2.11.6), we consider two reference frames S_a, S_b moving along the z-direction with velocities v_a, v_b with respect to our fixed frame S, and we assume that $\theta = 0^{\text{o}}$ in the frame S. Let f_a, f_b be the frequencies of the wave as measured in the frames S_a, S_b. Then, the separate application of Eq. (2.11.13) to S_a and S_b gives:

$$f_a = f\sqrt{\frac{1-\beta_a}{1+\beta_a}}\,,\quad f_b = f\sqrt{\frac{1-\beta_b}{1+\beta_b}}\,,\quad\Rightarrow\quad f_b = f_a\sqrt{\frac{1-\beta_b}{1+\beta_b}\cdot\frac{1+\beta_a}{1-\beta_a}} \tag{2.11.14}$$

where $\beta_a = v_a/c$ and $\beta_b = v_b/c$. This is equivalent to Eq. (2.11.6). The case when the wave is propagating in an arbitrary direction θ is given in Problem 2.27.

Next, we consider the transformation of the electromagnetic field components between the two frames. The electric field has the following form in S and S':

$$\mathbf{E} = \mathbf{E}_0 e^{j(\omega t - k_x x - k_z z)}\,,\quad \mathbf{E}' = \mathbf{E}_0' e^{j(\omega' t' - k_x' x' - k_z' z')} \tag{2.11.15}$$

As we discussed in Appendix H, the propagation phase factors remain invariant in the two frames, that is, $\omega t - k_x x - k_z z = \omega' t' - k_x' x' - k_z' z'$. Assuming a TE wave and using Eq. (2.9.9), the electric and magnetic field amplitudes will have the following form in the two frames:

$$\begin{aligned}
&\mathbf{E}_0 = E_0\hat{\mathbf{y}}\,,\quad c\mathbf{B}_0 = \eta_0\mathbf{H}_0 = \hat{\mathbf{k}}\times\mathbf{E}_0 = E_0(-\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{z}}\sin\theta)\\
&\mathbf{E}_0' = E_0'\hat{\mathbf{y}}\,,\quad c\mathbf{B}_0' = \eta_0\mathbf{H}_0' = \hat{\mathbf{k}}'\times\mathbf{E}_0' = E_0'(-\hat{\mathbf{x}}\cos\theta' + \hat{\mathbf{z}}\sin\theta')
\end{aligned} \tag{2.11.16}$$

Applying the Lorentz transformation properties of Eq. (H.31) to the above field components, we find:

$$\begin{aligned}
E_y' &= \gamma(E_y + \beta c B_x) & & E_0' = E_0\gamma(1-\beta\cos\theta)\\
cB_x' &= \gamma(cB_x + \beta E_y) & \Rightarrow\quad & -E_0'\cos\theta' = -E_0\gamma(\cos\theta - \beta)\\
cB_z' &= cB_z & & E_0'\sin\theta' = E_0\sin\theta
\end{aligned} \tag{2.11.17}$$

The first equation gives the desired relationship between E_0 and E_0'. The last two equations imply the same angle relationships as Eq. (2.11.11). The same relationship between E_0, E_0' holds also for a TM wave defined by $\mathbf{E}_0 = E_0(\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta)$.

2.12 Propagation in Negative-Index Media

In media with simultaneously negative permittivity and permeability, $\epsilon < 0$ and $\mu < 0$, the refractive index must be negative [353]. To see this, we consider a uniform plane wave propagating in a lossless medium:

$$E_x(z,t) = E_0\, e^{j\omega t - jkz}\,,\quad H_y(z,t) = H_0\, e^{j\omega t - jkz}$$

Then, Maxwell's equations require the following relationships, which are equivalent to Faraday's and Ampère's laws, respectively:

$$kE_0 = \omega\mu H_0\,,\quad kH_0 = \omega\epsilon E_0\,,\quad \text{or,}$$

$$\eta = \frac{E_0}{H_0} = \frac{\omega\mu}{k} = \frac{k}{\omega\epsilon}\quad\Rightarrow\quad k^2 = \omega^2\epsilon\mu \tag{2.12.1}$$

Because the medium is lossless, k and η will be real and the time-averaged Poynting vector, which points in the z-direction, will be:

$$\mathcal{P}_z = \frac{1}{2}\text{Re}[E_0 H_0^*] = \frac{1}{2\eta}|E_0|^2 = \frac{1}{2}\eta|H_0|^2 \tag{2.12.2}$$

If we require that the energy flux is towards the positive z-direction, that is, $\mathcal{P}_z > 0$, then we must have $\eta > 0$. Because μ and ϵ are negative, Eq. (2.12.1) implies that k must be negative, $k < 0$, in order for the ratio $\eta = \omega\mu/k$ to be positive. Thus, in solving $k^2 = \omega^2\mu\epsilon$, we must choose the negative square root:

$$k = -\omega\sqrt{\mu\epsilon} \tag{2.12.3}$$

The refractive index n may be defined through $k = k_0 n$, where $k_0 = \omega\sqrt{\mu_0\epsilon_0}$ is the free-space wavenumber. Thus, we have $n = k/k_0 = -\sqrt{\mu\epsilon/\mu_0\epsilon_0} = -\sqrt{\mu_{\text{rel}}\epsilon_{\text{rel}}}$, expressed in terms of the relative permittivity and permeability. Writing $\epsilon = -|\epsilon|$ and $\mu = -|\mu|$, we have for the medium impedance:

$$\eta = \frac{\omega\mu}{k} = \frac{-\omega|\mu|}{-\omega\sqrt{|\mu\epsilon|}} = \sqrt{\frac{|\mu|}{|\epsilon|}} = \sqrt{\frac{\mu}{\epsilon}} \tag{2.12.4}$$

which can be written also as follows, where $\eta_0 = \sqrt{\mu_0/\epsilon_0}$:

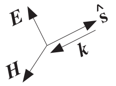
$$\eta = \eta_0\frac{\mu}{\mu_0 n} = \eta_0\frac{\epsilon_0 n}{\epsilon} \tag{2.12.5}$$

Thus, in negative-index media, the wave vector k and the phase velocity $v_{\text{ph}} = \omega/k = c_0/n$ will be negative, pointing in opposite direction than the Poynting vector. As we saw in Sec. 1.11, for lossless negative-index media the energy transport velocity v_{en}, which is in the direction of the Poynting vector, coincides with the group velocity v_g. Thus, $v_g = v_{\text{en}} > 0$, while $v_{\text{ph}} < 0$.

Two consequences of the negative refractive index, $n < 0$, are the reversal of Snel's law discussed in Sec. 7.16 and the possibility of a perfect lens discussed in Sec. 8.6. These and other consequences of $n < 0$, such as the reversal of the Doppler and Cherenkov effects and the reversal of the field momentum, have been discussed by Veselago [353].

If the propagation is along an arbitrary direction defined by a unit-vector $\hat{\mathbf{s}}$ (i.e., a rotated version of $\hat{\mathbf{z}}$), then we may define the wavevector by $\boldsymbol{k} = k\hat{\mathbf{s}}$, with k to be determined, and look for solutions of Maxwell's equations of the form:

$$\begin{aligned} \boldsymbol{E}(\mathbf{r}, t) &= \boldsymbol{E}_0\, e^{j\omega t - j\,\boldsymbol{k}\cdot\mathbf{r}} \\ \boldsymbol{H}(\mathbf{r}, t) &= \boldsymbol{H}_0\, e^{j\omega t - j\,\boldsymbol{k}\cdot\mathbf{r}} \end{aligned} \tag{2.12.6}$$



Gauss's laws require that the constant vectors $\boldsymbol{E}_0$, $\boldsymbol{H}_0$ be transverse to $\boldsymbol{k}$, or $\hat{\mathbf{s}}$, that is, $\hat{\mathbf{s}}\cdot\boldsymbol{E}_0 = \hat{\mathbf{s}}\cdot\boldsymbol{H}_0 = 0$. Then, Faraday's and Ampère's laws require that:

$$\boldsymbol{H}_0 = \frac{1}{\eta}\,(\hat{\mathbf{s}}\times\boldsymbol{E}_0)\,, \quad \eta = \frac{\omega\mu}{k} = \frac{k}{\omega\epsilon} \quad \Rightarrow \quad k^2 = \omega^2\mu\epsilon \tag{2.12.7}$$

with a Poynting vector:

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\,\mathrm{Re}\big[\boldsymbol{E}_0\times\boldsymbol{H}_0^*\big] = \hat{\mathbf{s}}\,\frac{1}{2\eta}\,|\boldsymbol{E}_0|^2 \tag{2.12.8}$$

Thus, if $\boldsymbol{\mathcal{P}}$ is assumed to be in the direction of $\hat{\mathbf{s}}$, then we must have $\eta > 0$, and therefore, k must be negative as in Eq. (2.12.3). It follows that the wavevector $\boldsymbol{k} = k\hat{\mathbf{s}}$ will be in the opposite direction of $\hat{\mathbf{s}}$ and $\boldsymbol{\mathcal{P}}$. Eq. (2.12.7) implies that the triplet $\{\boldsymbol{E}_0, \boldsymbol{H}_0, \hat{\mathbf{s}}\}$ is still a right-handed vector system, but $\{\boldsymbol{E}_0, \boldsymbol{H}_0, \boldsymbol{k}\}$ will be a *left-handed* system. This is the reason why Veselago [353] named such media *left-handed media.*[†]

In a lossy negative-index medium, the permittivity and permeability will be complex-valued, $\epsilon = \epsilon_r - j\epsilon_i$ and $\mu = \mu_r - j\mu_i$, with negative real parts $\epsilon_r, \mu_r < 0$, and positive imaginary parts $\epsilon_i, \mu_i > 0$. Eq. (2.12.1) remains the same and will imply that k and η will be complex-valued. Letting $k = \beta - j\alpha$, the fields will be attenuating as they propagate:

$$E_x(z,t) = E_0 e^{-\alpha z} e^{j\omega t - j\beta z}\,, \quad H_y(z,t) = H_0 e^{-\alpha z} e^{j\omega t - j\beta z}$$

and the Poynting vector will be given by:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}\big[E_x(z) H_y^*(z)\big] = \frac{1}{2}\,\mathrm{Re}\Big(\frac{1}{\eta}\Big)|E_0|^2 e^{-2\alpha z} = \frac{1}{2}\,\mathrm{Re}(\eta)\,|H_0|^2 e^{-2\alpha z} \tag{2.12.9}$$

The refractive index is complex-valued, $n = n_r - jn_i$, and is related to k through $k = k_0 n$, or, $\beta - j\alpha = k_0(n_r - jn_i)$, or, $\beta = k_0 n_r$ and $\alpha = k_0 n_i$. Thus, the conditions of negative phase velocity ($\beta < 0$), field attenuation ($\alpha > 0$), and positive power flow can be stated equivalently as follows:

$$n_r < 0\,, \quad n_i > 0\,, \quad \mathrm{Re}(\eta) > 0 \tag{2.12.10}$$

Next, we look at the necessary and sufficient conditions for a medium to satisfy these conditions. If we express ϵ, μ in their polar forms, $\epsilon = |\epsilon| e^{-j\theta_\epsilon}$ and $\mu = |\mu| e^{-j\theta_\mu}$, then, regardless of the signs of ϵ_r, μ_r, the assumption that the medium is lossy, $\epsilon_i, \mu_i > 0$, requires that $\sin\theta_\epsilon > 0$ and $\sin\theta_\mu > 0$, and these are equivalent to the restrictions:

$$0 \le \theta_\epsilon \le \pi\,, \quad 0 \le \theta_\mu \le \pi \tag{2.12.11}$$

[†] The term negative-index media is preferred in order to avoid confusion with chiral media.

To guarantee $\alpha > 0$, the wavenumber k must be computed by taking the positive square root of $k^2 = \omega^2\mu\epsilon = \omega^2|\mu\epsilon|^2 e^{-j(\theta_\epsilon + \theta_\mu)}$, that is,

$$k = \beta - j\alpha = \omega\sqrt{|\mu\epsilon|}\,e^{-j\theta_+}\,, \quad \theta_+ = \frac{\theta_\epsilon + \theta_\mu}{2} \tag{2.12.12}$$

Indeed, the restrictions (2.12.11) imply the same for θ_+, that is, $0 \le \theta_+ \le \pi$, or, equivalently, $\sin\theta_+ > 0$, and hence $\alpha > 0$. Similarly, the quantities n, η are given by:

$$n = |n|\,e^{-j\theta_+}\,, \quad \eta = |\eta|\,e^{-j\theta_-}\,, \quad \theta_- = \frac{\theta_\epsilon - \theta_\mu}{2} \tag{2.12.13}$$

where $|n| = \sqrt{|\mu\epsilon|/\mu_0\epsilon_0}$ and $|\eta| = \sqrt{|\mu|/|\epsilon|}$. It follows that $n_i = |n|\sin\theta_+ > 0$. Since $n_r = |n|\cos\theta_+$ and $\mathrm{Re}(\eta) = |\eta|\cos\theta_-$, the conditions $n_r < 0$ and $\mathrm{Re}(\eta) > 0$ will be equivalent to

$$\cos\theta_+ = \cos\Big(\frac{\theta_\epsilon + \theta_\mu}{2}\Big) < 0\,, \quad \cos\theta_- = \cos\Big(\frac{\theta_\epsilon - \theta_\mu}{2}\Big) > 0 \tag{2.12.14}$$

Using some trigonometric identities, these conditions become equivalently:

$$\begin{aligned} \cos(\theta_\epsilon/2)\cos(\theta_\mu/2) - \sin(\theta_\epsilon/2)\sin(\theta_\mu/2) &< 0 \\ \cos(\theta_\epsilon/2)\cos(\theta_\mu/2) + \sin(\theta_\epsilon/2)\sin(\theta_\mu/2) &> 0 \end{aligned}$$

which combine into

$$-\sin(\theta_\epsilon/2)\sin(\theta_\mu/2) < \cos(\theta_\epsilon/2)\cos(\theta_\mu/2) < \sin(\theta_\epsilon/2)\sin(\theta_\mu/2)$$

Because $0 \le \theta_\epsilon/2 \le \pi/2$, we have $\cos(\theta_\epsilon/2) \ge 0$ and $\sin(\theta_\epsilon/2) \ge 0$, and similarly for $\theta_\mu/2$. Thus, the above conditions can be replaced by the single equivalent inequality:

$$\tan(\theta_\epsilon/2)\tan(\theta_\mu/2) > 1 \tag{2.12.15}$$

A number of equivalent conditions have been given in the literature [374,402] for a medium to have negative phase velocity and positive power:

$$\begin{aligned} (|\epsilon| - \epsilon_r)(|\mu| - \mu_r) &> \epsilon_i\mu_i \\ \epsilon_r|\mu| + \mu_r|\epsilon| &< 0 \\ \epsilon_r\mu_i + \mu_r\epsilon_i &< 0 \end{aligned} \tag{2.12.16}$$

They are all equivalent to condition (2.12.15). This can be seen by writing them in terms of the angles $\theta_\epsilon, \theta_\mu$ and then using simple trigonometric identities, such as $\tan(\theta/2) = (1 - \cos\theta)/\sin\theta$, to show their equivalence to (2.12.15):

$$\begin{aligned} (1 - \cos\theta_\epsilon)(1 - \cos\theta_\mu) &> \sin\theta_\epsilon\sin\theta_\mu \\ \cos\theta_\epsilon + \cos\theta_\mu &< 0 \\ \cot\theta_\epsilon + \cot\theta_\mu &< 0 \end{aligned} \tag{2.12.17}$$

If the medium has negative real parts, $\epsilon_r < 0$ and $\mu_r < 0$, then the conditions (2.12.16) are obviously satisfied.

2.13 Problems

2.1 A function $E(z,t)$ may be thought of as a function $E(\zeta,\xi)$ of the independent variables $\zeta = z - ct$ and $\xi = z + ct$. Show that the wave equation (2.1.6) and the forward-backward equations (2.1.10) become in these variables:

$$\frac{\partial^2 \boldsymbol{E}}{\partial\zeta\,\partial\xi}=0\,,\qquad \frac{\partial \boldsymbol{E}_+}{\partial\xi}=0\,,\qquad \frac{\partial \boldsymbol{E}_-}{\partial\zeta}=0$$

Thus, $\boldsymbol{E}_+$ may depend only on ζ and $\boldsymbol{E}_-$ only on ξ.

2.2 A source located at $z = 0$ generates an electromagnetic pulse of duration of T sec, given by $\boldsymbol{E}(0,t)= \hat{\mathbf{x}}E_0\,\big[u(t)-u(t-T)\big]$, where $u(t)$ is the unit step function and E_0 is a constant. The pulse is launched towards the positive z-direction. Determine expressions for $\boldsymbol{E}(z,t)$ and $\boldsymbol{H}(z,t)$ and sketch them versus z at any given t.

2.3 Show that for a single-frequency wave propagating along the z-direction the corresponding transverse fields $\boldsymbol{E}(z),\boldsymbol{H}(z)$ satisfy the system of equations:

$$\frac{\partial}{\partial z}\begin{bmatrix}\boldsymbol{E}\\ \boldsymbol{H}\times\hat{\mathbf{z}}\end{bmatrix}=\begin{bmatrix}0 & -j\omega\mu\\ -j\omega\epsilon & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{E}\\ \boldsymbol{H}\times\hat{\mathbf{z}}\end{bmatrix}$$

where the matrix equation is meant to apply individually to the x,y components of the vector entries. Show that the following similarity transformation diagonalizes the transition matrix, and discuss its role in decoupling and solving the above system in terms of forward and backward waves:

$$\begin{bmatrix}1 & \eta\\ 1 & -\eta\end{bmatrix}\begin{bmatrix}0 & -j\omega\mu\\ -j\omega\epsilon & 0\end{bmatrix}\begin{bmatrix}1 & \eta\\ 1 & -\eta\end{bmatrix}^{-1}=\begin{bmatrix}-jk & 0\\ 0 & jk\end{bmatrix}$$

where $k=\omega/c$, $c=1/\sqrt{\mu\epsilon}$, and $\eta=\sqrt{\mu/\epsilon}$.

2.4 The visible spectrum has the wavelength range 380–780 nm. What is this range in THz? In particular, determine the frequencies of red, orange, yellow, green, blue, and violet having the nominal wavelengths of 700, 610, 590, 530, 470, and 420 nm.

2.5 What is the frequency in THz of a typical $\mathrm{CO_2}$ laser (used in laser surgery) having the far infrared wavelength of 20 μm?

2.6 What is the wavelength in meters or cm of a wave with the frequencies of 10 kHz, 10 MHz, and 10 GHz?

What is the frequency in GHz of the 21-cm hydrogen line observed in the cosmos?

What is the wavelength in cm of the typical microwave oven frequency of 2.45 GHz?

2.7 Suppose you start with $\boldsymbol{E}(z,t)=\hat{\mathbf{x}}E_0e^{j\omega t-jkz}$, but you do not yet know the relationship between k and ω (you may assume they are both positive.) By inserting $\boldsymbol{E}(z,t)$ into Maxwell's equations, determine the k–ω relationship as a consequence of these equations. Determine also the magnetic field $\boldsymbol{H}(z,t)$ and verify that all of Maxwell's equations are satisfied.

Repeat the problem if $\boldsymbol{E}(z,t)=\hat{\mathbf{x}}E_0e^{j\omega t+jkz}$ and if $\boldsymbol{E}(z,t)=\hat{\mathbf{y}}E_0e^{j\omega t-jkz}$.

2.8 Determine the polarization types of the following waves, and indicate the direction, if linear, and the sense of rotation, if circular or elliptic:

a. $\boldsymbol{E}=E_0(\hat{\mathbf{x}}+\hat{\mathbf{y}})e^{-jkz}$
b. $\boldsymbol{E}=E_0(\hat{\mathbf{x}}-\sqrt{3}\,\hat{\mathbf{y}})e^{-jkz}$
c. $\boldsymbol{E}=E_0(j\hat{\mathbf{x}}+\hat{\mathbf{y}})e^{-jkz}$
d. $\boldsymbol{E}=E_0(\hat{\mathbf{x}}-2j\hat{\mathbf{y}})e^{-jkz}$
e. $\boldsymbol{E}=E_0(\hat{\mathbf{x}}-\hat{\mathbf{y}})e^{-jkz}$
f. $\boldsymbol{E}=E_0(\sqrt{3}\,\hat{\mathbf{x}}-\hat{\mathbf{y}})e^{-jkz}$
g. $\boldsymbol{E}=E_0(j\hat{\mathbf{x}}-\hat{\mathbf{y}})e^{jkz}$
h. $\boldsymbol{E}=E_0(\hat{\mathbf{x}}+2j\hat{\mathbf{y}})e^{jkz}$

2.9 A uniform plane wave, propagating in the z-direction in vacuum, has the following electric field:

$$\boldsymbol{\mathcal{E}}(t,z)=2\,\hat{\mathbf{x}}\cos(\omega t-kz)+4\,\hat{\mathbf{y}}\sin(\omega t-kz)$$

a. Determine the vector phasor representing $\boldsymbol{\mathcal{E}}(t,z)$ in the complex form $\boldsymbol{E}=\boldsymbol{E}_0e^{j\omega t-jkz}$.

b. Determine the polarization of this electric field (linear, circular, elliptic, left-handed, right-handed?)

c. Determine the magnetic field $\boldsymbol{\mathcal{H}}(t,z)$ in its real-valued form.

2.10 A uniform plane wave propagating in vacuum along the z direction has real-valued electric field components:

$$\mathcal{E}_x(z,t)=\cos(\omega t-kz)\,,\quad \mathcal{E}_y(z,t)=2\sin(\omega t-kz)$$

a. Its phasor form has the form $\boldsymbol{E}=(A\,\hat{\mathbf{x}}+B\,\hat{\mathbf{y}})e^{\pm jkz}$. Determine the numerical values of the complex-valued coefficients A,B and the correct sign of the exponent.

b. Determine the polarization of this wave (left, right, linear, etc.). Explain your reasoning.

2.11 Consider the two electric fields, one given in its real-valued form, and the other, in its phasor form:

a. $\boldsymbol{\mathcal{E}}(t,z)=\hat{\mathbf{x}}\sin(\omega t+kz)+2\hat{\mathbf{y}}\cos(\omega t+kz)$
b. $\boldsymbol{E}(z)=\big[(1+j)\hat{\mathbf{x}}-(1-j)\hat{\mathbf{y}}\big]e^{-jkz}$

For both cases, determine the polarization of the wave (linear, circular, left, right, etc.) and the direction of propagation.

For case (a), determine the field in its phasor form. For case (b), determine the field in its real-valued form as a function of t,z.

2.12 A uniform plane wave propagating in the z-direction has the following real-valued electric field:

$$\boldsymbol{\mathcal{E}}(t,z)=\hat{\mathbf{x}}\cos(\omega t-kz-\pi/4)+\hat{\mathbf{y}}\cos(\omega t-kz+\pi/4)$$

a. Determine the complex-phasor form of this electric field.

b. Determine the corresponding magnetic field $\boldsymbol{\mathcal{H}}(t,z)$ given in its real-valued form.

c. Determine the polarization type (left, right, linear, etc.) of this wave.

2.13 Determine the polarization type (left, right, linear, etc.) and the direction of propagation of the following electric fields given in their phasor forms:

a. $\boldsymbol{E}(z)=\big[(1+j\sqrt{3})\hat{\mathbf{x}}+2\,\hat{\mathbf{y}}\big]e^{+jkz}$
b. $\boldsymbol{E}(z)=\big[(1+j)\hat{\mathbf{x}}-(1-j)\hat{\mathbf{y}}\big]e^{-jkz}$
c. $\boldsymbol{E}(z)=\big[\hat{\mathbf{x}}-\hat{\mathbf{z}}+j\sqrt{2}\,\hat{\mathbf{y}}\big]e^{-jk(x+z)/\sqrt{2}}$

2.14 Consider a forward-moving wave in its real-valued form:

$$\boldsymbol{\mathcal{E}}(t,z)=\hat{\mathbf{x}}A\cos(\omega t-kz+\phi_a)+\hat{\mathbf{y}}B\cos(\omega t-kz+\phi_b)$$

Show that:

$$\boldsymbol{\mathcal{E}}(t+\Delta t,z+\Delta z)\times\boldsymbol{\mathcal{E}}(t,z)=\hat{\mathbf{z}}\,AB\sin(\phi_a-\phi_b)\sin(\omega\Delta t-k\Delta z)$$

2.15 Show that in order for the polarization ellipse of Eq. (2.5.4) to be equivalent to the rotated one of Eq. (2.5.7), one must determine the tilt angle θ such that the following matrix condition is satisfied:

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \dfrac{1}{A^2} & -\dfrac{\cos\phi}{AB} \\ -\dfrac{\cos\phi}{AB} & \dfrac{1}{B^2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \sin^2\phi \begin{bmatrix} \dfrac{1}{A'^2} & 0 \\ 0 & \dfrac{1}{B'^2} \end{bmatrix}$$

Show that the required angle θ is given by Eq. (2.5.5). Then, show that the following condition is satisfied, where $\tau = \tan\theta$:

$$\frac{(A^2 - B^2\tau^2)\,(B^2 - A^2\tau^2)}{(1-\tau^2)^2} = A^2B^2\sin^2\phi$$

Using this property, show that the semi-axes A', B' are given by the equations:

$$A'^2 = \frac{A^2 - B^2\tau^2}{1-\tau^2}\,, \qquad B'^2 = \frac{B^2 - A^2\tau^2}{1-\tau^2}$$

Then, transform these equations into the form of Eq. (2.5.6). Finally, show that A', B' satisfy the relationships:

$$A'^2 + B'^2 = A^2 + B^2\,, \quad A'B' = AB|\sin\phi|$$

2.16 Show the cross-product equation (2.5.10). Then, prove the more general relationship:

$$\boldsymbol{\mathcal{E}}(t_1)\times\boldsymbol{\mathcal{E}}(t_2) = \hat{\mathbf{z}}\,AB\sin\phi\sin\big(\omega\,(t_2 - t_1)\big)$$

Discuss how linear polarization can be explained with the help of this result.

2.17 Using the properties $k_c\eta_c = \omega\mu$ and $k_c^2 = \omega^2\mu\epsilon_c$ for the complex-valued quantities k_c, η_c of Eq. (2.6.5), show the following relationships, where $\epsilon_c = \epsilon' - j\epsilon''$ and $k_c = \beta - j\alpha$:

$$\mathrm{Re}\big(\eta_c^{-1}\big) = \frac{\omega\epsilon''}{2\alpha} = \frac{\beta}{\omega\mu}$$

2.18 Show that for a lossy medium the complex-valued quantities k_c and η_c may be expressed as follows, in terms of the loss angle θ defined in Eq. (2.6.27):

$$k_c = \beta - j\alpha = \omega\sqrt{\mu\epsilon_d'}\left(\cos\frac{\theta}{2} - j\sin\frac{\theta}{2}\right)(\cos\theta)^{-1/2}$$

$$\eta_c = \eta' + j\eta'' = \sqrt{\frac{\mu}{\epsilon_d'}}\left(\cos\frac{\theta}{2} + j\sin\frac{\theta}{2}\right)(\cos\theta)^{1/2}$$

2.19 It is desired to reheat frozen mashed potatoes and frozen cooked carrots in a microwave oven operating at 2.45 GHz. Determine the penetration depth and assess the effectiveness of this heating method. Moreover, determine the attenuation of the electric field (in dB and absolute units) at a depth of 1 cm from the surface of the food. The complex dielectric constants of the mashed potatoes and carrots are (see [114]) $\epsilon_c = (65 - j25)\epsilon_0$ and $\epsilon_c = (75 - j25)\epsilon_0$.

2.20 We wish to shield a piece of equipment from RF interference over the frequency range from 10 kHz to 1 GHz by enclosing it in a copper enclosure. The RF interference inside the enclosure is required to be at least 50 dB down compared to its value outside. What is the minimum thickness of the copper shield in mm?

2.21 In order to protect a piece of equipment from RF interference, we construct an enclosure made of aluminum foil (you may assume a reasonable value for its thickness.) The conductivity of aluminum is 3.5×10^7 S/m. Over what frequency range can this shield protect our equipment assuming the same 50-dB attenuation requirement of the previous problem?

2.22 A uniform plane wave propagating towards the positive z-direction in empty space has an electric field at $z = 0$ that is a linear superposition of two components of frequencies ω_1 and ω_2:

$$\boldsymbol{E}(0,t) = \hat{\mathbf{x}}\,(E_1e^{j\omega_1 t} + E_2e^{j\omega_2 t})$$

Determine the fields $\boldsymbol{E}(z,t)$ and $\boldsymbol{H}(z,t)$ at any point z.

2.23 An electromagnetic wave propagating in a lossless dielectric is described by the electric and magnetic fields, $\boldsymbol{E}(z) = \hat{\mathbf{x}}\,E(z)$ and $\boldsymbol{H}(z) = \hat{\mathbf{y}}\,H(z)$, consisting of the forward and backward components:

$$E(z) = E_+e^{-jkz} + E_-e^{jkz}$$

$$H(z) = \frac{1}{\eta}(E_+e^{-jkz} - E_-e^{jkz})$$

a. Verify that these expressions satisfy all of Maxwell's equations.

b. Show that the time-averaged energy flux in the z-direction is independent of z and is given by:

$$\mathcal{P}_z = \frac{1}{2}\mathrm{Re}\big[E(z)H^*(z)\big] = \frac{1}{2\eta}\,(|E_+|^2 - |E_-|^2)$$

c. Assuming $\mu = \mu_0$ and $\epsilon = n^2\epsilon_0$, so that n is the refractive index of the dielectric, show that the fields at two different z-locations, say at $z = z_1$ and $z = z_2$ are related by the matrix equation:

$$\begin{bmatrix} E(z_1) \\ \eta_0 H(z_1) \end{bmatrix} = \begin{bmatrix} \cos kl & jn^{-1}\sin kl \\ jn\sin kl & \cos kl \end{bmatrix} \begin{bmatrix} E(z_2) \\ \eta_0 H(z_2) \end{bmatrix}$$

where $l = z_2 - z_1$, and we multiplied the magnetic field by $\eta_0 = \sqrt{\mu_0/\epsilon_0}$ in order to give it the same dimensions as the electric field.

d. Let $Z(z) = \dfrac{E(z)}{\eta_0 H(z)}$ and $Y(z) = \dfrac{1}{Z(z)}$ be the *normalized* wave impedance and admittance at location z. Show the relationships at at the locations z_1 and z_2:

$$Z(z_1) = \frac{Z(z_2) + jn^{-1}\tan kl}{1 + jnZ(z_2)\tan kl}\,, \qquad Y(z_1) = \frac{Y(z_2) + jn\tan kl}{1 + jn^{-1}Y(z_2)\tan kl}$$

What would be these relationships if had we normalized to the medium impedance, that is, $Z(z) = E(z)/\eta H(z)$?

2.24 Show that the time-averaged energy density and Poynting vector of the obliquely moving wave of Eq. (2.9.10) are given by

$$w = \frac{1}{2}\mathrm{Re}\Big[\frac{1}{2}\epsilon\boldsymbol{E}\cdot\boldsymbol{E}^* + \frac{1}{2}\mu\boldsymbol{H}\cdot\boldsymbol{H}^*\Big] = \frac{1}{2}\epsilon\big(|A|^2 + |B|^2\big)$$

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\mathrm{Re}\big[\boldsymbol{E}\times\boldsymbol{H}^*\big] = \hat{\mathbf{z}}'\,\frac{1}{2\eta}\big(|A|^2 + |B|^2\big) = (\hat{\mathbf{z}}\cos\theta + \hat{\mathbf{x}}\sin\theta)\,\frac{1}{2\eta}\big(|A|^2 + |B|^2\big)$$

where $\hat{\mathbf{z}}' = \hat{\mathbf{z}}\cos\theta + \hat{\mathbf{x}}\sin\theta$ is the unit vector in the direction of propagation. Show that the energy transport velocity is $\boldsymbol{v} = \boldsymbol{\mathcal{P}}/w = c\,\hat{\mathbf{z}}'$.

2.25 A uniform plane wave propagating in empty space has electric field:

$$\mathbf{E}(x,z,t)=\hat{\mathbf{y}}\,E_0 e^{j\omega t}e^{-jk(x+z)/\sqrt{2}}\,,\qquad k=\frac{\omega}{c_0}$$

a. Inserting $\mathbf{E}(x,z,t)$ into Maxwell's equations, work out an expression for the corresponding magnetic field $\mathbf{H}(x,z,t)$.

b. What is the direction of propagation and its unit vector $\hat{\mathbf{k}}$?

c. Working with Maxwell's equations, determine the electric field $\mathbf{E}(x,z,t)$ and propagation direction $\hat{\mathbf{k}}$, if we started with a magnetic field given by:

$$\mathbf{H}(x,z,t)=\hat{\mathbf{y}}\,H_0 e^{j\omega t}e^{-jk(\sqrt{3}z-x)/2}$$

2.26 A linearly polarized light wave with electric field $\mathbf{E}_0$ at angle θ with respect to the x-axis is incident on a polarizing filter, followed by an identical polarizer (the analyzer) whose primary axes are rotated by an angle ϕ relative to the axes of the first polarizer, as shown in Fig. 2.13.1.

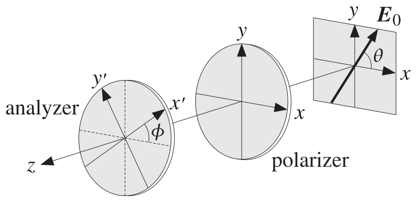


Fig. 2.13.1 Polarizer–analyzer filter combination.

Assume that the amplitude attenuations through the first polarizer are a_1, a_2 with respect to the x- and y-directions. The polarizer transmits primarily the x-polarization, so that $a_2 \ll a_1$. The analyzer is rotated by an angle ϕ so that the same gains a_1, a_2 now refer to the x'- and y'-directions.

a. Ignoring the phase retardance introduced by each polarizer, show that the polarization vectors at the input, and after the first and second polarizers, are:

$$\begin{aligned}
\mathbf{E}_0 &= \hat{\mathbf{x}}\cos\theta+\hat{\mathbf{y}}\sin\theta\\
\mathbf{E}_1 &= \hat{\mathbf{x}}\,a_1\cos\theta+\hat{\mathbf{y}}a_2\sin\theta\\
\mathbf{E}_2 &= \hat{\mathbf{x}}'\,(a_1^2\cos\phi\cos\theta+a_1a_2\sin\phi\sin\theta)+\hat{\mathbf{y}}'\,(a_2^2\cos\phi\sin\theta-a_1a_2\sin\phi\cos\theta)
\end{aligned}$$

where $\{\hat{\mathbf{x}}',\hat{\mathbf{y}}'\}$ are related to $\{\hat{\mathbf{x}},\hat{\mathbf{y}}\}$ as in Problem 4.7.

b. Explain the meaning and usefulness of the matrix operations:

$$\begin{bmatrix} a_1 & 0\\ 0 & a_2\end{bmatrix}\begin{bmatrix}\cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix} a_1 & 0\\ 0 & a_2\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}\qquad\text{and}$$

$$\begin{bmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix} a_1 & 0\\ 0 & a_2\end{bmatrix}\begin{bmatrix}\cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix} a_1 & 0\\ 0 & a_2\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$$

c. Show that the output light intensity is proportional to the quantity:

$$\begin{aligned}
I = &(a_1^4\cos^2\theta+a_2^4\sin^2\theta)\cos^2\phi+a_1^2a_2^2\sin^2\phi+\\
&+2a_1a_2(a_1^2-a_2^2)\cos\phi\sin\phi\cos\theta\sin\theta
\end{aligned}$$

d. If the input light were unpolarized, that is, incoherent, show that the average of the intensity of part (c) over all angles $0\le\theta\le 2\pi$, will be given by the generalized Malus's law:

$$\bar{I}=\frac{1}{2}(a_1^4+a_2^4)\cos^2\phi+a_1^2a_2^2\sin^2\phi$$

The case $a_2=0$, represents the usual *Malus' law.*

2.27 First, prove the equivalence of the three relationships given in Eq. (2.11.11). Then, prove the following identity between the angles θ,θ':

$$(1-\beta\cos\theta)\,(1+\beta\cos\theta')=(1+\beta\cos\theta)\,(1-\beta\cos\theta')=1-\beta^2$$

Using this identity, prove the alternative Doppler formulas (2.11.12).

2.28 In proving the relativistic Doppler formula (2.11.14), it was assumed that the plane wave was propagating in the z-direction in all three reference frames S, S_a, S_b. If in the frame S the wave is propagating along the θ-direction shown in Fig. 2.11.2, show that the Doppler formula may be written in the following equivalent forms:

$$f_b=f_a\frac{\gamma_b(1-\beta_b\cos\theta)}{\gamma_a(1-\beta_a\cos\theta)}=f_a\gamma(1-\beta\cos\theta_a)=\frac{f_a}{\gamma(1+\beta\cos\theta_b)}=f_a\sqrt{\frac{1-\beta\cos\theta_a}{1+\beta\cos\theta_b}}$$

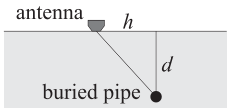
where

$$\beta_a=\frac{v_a}{c}\,,\quad\beta_b=\frac{v_b}{c}\,,\quad\beta=\frac{v}{c}\,,\quad\gamma_a=\frac{1}{\sqrt{1-\beta_a^2}}\,,\quad\gamma_b=\frac{1}{\sqrt{1-\beta_b^2}}\,,\quad\gamma=\frac{1}{\sqrt{1-\beta^2}}$$

and v is the relative velocity of the observer and source given by Eq. (2.11.7), and θ_a, θ_b are the propagation directions in the frames S_a, S_b. Moreover, show the following relations among these angles:

$$\cos\theta_a=\frac{\cos\theta-\beta_a}{1-\beta_a\cos\theta}\,,\quad\cos\theta_b=\frac{\cos\theta-\beta_b}{1-\beta_b\cos\theta}\,,\quad\cos\theta_b=\frac{\cos\theta_a-\beta}{1-\beta\cos\theta_a}$$

2.29 Ground-penetrating radar operating at 900 MHz is used to detect underground objects, as shown in the figure below for a buried pipe. Assume that the earth has conductivity $\sigma = 10^{-3}$ S/m, permittivity $\epsilon=9\epsilon_0$, and permeability $\mu=\mu_0$. You may use the "weakly lossy dielectric" approximation.



a. Determine the numerical value of the wavenumber $k=\beta-j\alpha$ in meters^{-1}, and the penetration depth $\delta=1/\alpha$ in meters.

b. Determine the value of the complex refractive index $n_c=n_r-jn_i$ of the ground at 900 MHz.

c. With reference to the above figure, explain why the electric field returning back to the radar antenna after getting reflected by the buried pipe is given by

$$\left|\frac{E_{\text{ret}}}{E_0}\right|^2 = \exp\left[-\frac{4\sqrt{h^2+d^2}}{\delta}\right]$$

where E_0 is the transmitted signal, d is the depth of the pipe, and h is the horizontal displacement of the antenna from the pipe. You may ignore the angular response of the radar antenna and assume it emits isotropically in all directions into the ground.

d. The depth d may be determined by measuring the roundtrip time $t(h)$ of the transmitted signal at successive horizontal distances h. Show that $t(h)$ is given by:

$$t(h) = \frac{2n_r}{c_0}\sqrt{d^2+h^2}$$

where n_r is the real part of the complex refractive index n_c.

e. Suppose $t(h)$ is measured over the range $-2 \le h \le 2$ meters over the pipe and its minimum recorded value is $t_{\text{min}} = 0.2\ \mu$sec. What is the depth d in meters?

3

Pulse Propagation in Dispersive Media

In this chapter, we examine some aspects of pulse propagation in dispersive media and the role played by various wave velocity definitions, such as phase, group, and front velocities. We discuss group velocity dispersion, pulse spreading, chirping, and dispersion compensation, and look at some slow, fast, and negative group velocity examples. We also present a short introduction to chirp radar and pulse compression, elaborating on the similarities to dispersion compensation. The similarities to Fresnel diffraction and Fourier optics are discussed in Sec. 17.18. The chapter ends with a guide to the literature in these diverse topics.

3.1 Propagation Filter

As we saw in the previous chapter, a monochromatic plane wave moving forward along the z-direction has an electric field $E(z)=E(0)\,e^{-jkz}$, where $E(z)$ stands for either the x or the y component. We assume a homogeneous isotropic non-magnetic medium ($\mu=\mu_0$) with an effective permittivity $\epsilon(\omega)$; therefore, k is the frequency-dependent and possibly complex-valued wavenumber defined by $k(\omega)=\omega\sqrt{\epsilon(\omega)\mu_0}$. To emphasize the dependence on the frequency ω, we rewrite the propagated field as:[†]

$$\hat{E}(z,\omega)=e^{-jkz}\hat{E}(0,\omega) \tag{3.1.1}$$

Its complete space-time dependence will be:

$$e^{j\omega t}\hat{E}(z,\omega)=e^{j(\omega t-kz)}\hat{E}(0,\omega) \tag{3.1.2}$$

A wave packet or pulse can be made up by adding different frequency components, that is, by the inverse Fourier transform:

$$\boxed{E(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j(\omega t-kz)}\hat{E}(0,\omega)\,d\omega} \tag{3.1.3}$$

[†] where the hat denotes Fourier transformation.

Setting $z = 0$, we recognize $\hat{E}(0,\omega)$ to be the Fourier transform of the initial waveform $E(0,t)$, that is,

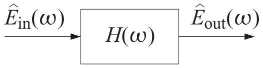
$$E(0,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}\hat{E}(0,\omega)\,d\omega \quad\Leftrightarrow\quad \hat{E}(0,\omega) = \int_{-\infty}^{\infty} e^{-j\omega t}E(0,t)\,dt \tag{3.1.4}$$

The multiplicative form of Eq. (3.1.1) allows us to think of the propagated field as the output of a *linear system*, the propagation filter, whose frequency response is

$$\boxed{H(z,\omega) = e^{-jk(\omega)z}} \tag{3.1.5}$$

Indeed, for a linear time-invariant system with impulse response $h(t)$ and corresponding frequency response $H(\omega)$, the input/output relationship can be expressed multiplicatively in the frequency domain or convolutionally in the time domain:

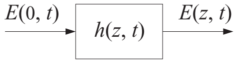
$$\begin{aligned}\hat{E}_{\text{out}}(\omega) &= H(\omega)\hat{E}_{\text{in}}(\omega)\\ E_{\text{out}}(t) &= \int_{-\infty}^{\infty} h(t-t')E_{\text{in}}(t')\,dt'\end{aligned}$$



For the propagator frequency response $H(z,\omega) = e^{-jk(\omega)z}$, we obtain the corresponding impulse response:

$$\boxed{h(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}H(z,\omega)\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega t - kz)}\,d\omega} \tag{3.1.6}$$

Alternatively, Eq. (3.1.6) follows from (3.1.3) by setting $\hat{E}(0,\omega) = 1$, corresponding to an impulsive input $E(0,t) = \delta(t)$. Thus, Eq. (3.1.3) may be expressed in the time domain in the convolutional form:



$$\boxed{E(z,t) = \int_{-\infty}^{\infty} h(z,t-t')E(0,t')\,dt'} \tag{3.1.7}$$

Example 3.1.1: For propagation in a dispersionless medium with frequency-independent permittivity, such as the vacuum, we have $k = \omega/c$, where $c = 1/\sqrt{\mu\epsilon}$. Therefore,

$$H(z,\omega) = e^{-jk(\omega)z} = e^{-j\omega z/c} = \text{pure delay by } z/c$$

$$h(z,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega t - kz)}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega(t - z/c)}\,d\omega = \delta(t - z/c)$$

and Eq. (3.1.7) gives $E(z,t) = E(0, t - z/c)$, in agreement with the results of Sec. 2.1. □

The reality of $h(z,t)$ implies the hermitian property, $H(z,-\omega)^* = H(z,\omega)$, for the frequency response, which is equivalent to the anti-hermitian property for the wavenumber, $k(-\omega)^* = -k(\omega)$.

3.2 *Front Velocity and Causality*

For a general linear system $H(\omega) = |H(\omega)|e^{-j\phi(\omega)}$, one has the standard concepts of phase delay, group delay, and signal-front delay [155] defined in terms of the system's phase-delay response, that is, the negative of its phase response, $\phi(\omega) = -\text{Arg}\,H(\omega)$:

$$t_p = \frac{\phi(\omega)}{\omega}\,,\quad t_g = \frac{d\phi(\omega)}{d\omega}\,,\quad t_f = \lim_{\omega\to\infty}\frac{\phi(\omega)}{\omega} \tag{3.2.1}$$

The significance of the signal-front delay t_f for the causality of a linear system is that the impulse response vanishes, $h(t) = 0$, for $t < t_f$, which implies that if the input begins at time $t = t_0$, then the output will begin at $t = t_0 + t_f$:

$$E_{\text{in}}(t) = 0 \;\text{ for }\; t < t_0 \quad\Rightarrow\quad E_{\text{out}}(t) = 0 \;\text{ for }\; t < t_0 + t_f \tag{3.2.2}$$

To apply these concepts to the propagator filter, we write $k(\omega)$ in terms of its real and imaginary parts, $k(\omega) = \beta(\omega) - j\alpha(\omega)$, so that

$$H(z,\omega) = e^{-jk(\omega)z} = e^{-\alpha(\omega)z}e^{-j\beta(\omega)z} \quad\Rightarrow\quad \phi(\omega) = \beta(\omega)z \tag{3.2.3}$$

Then, the definitions (3.2.1) lead naturally to the concepts of phase velocity, group velocity, and signal-front velocity, defined through:

$$t_p = \frac{z}{v_p}\,,\quad t_g = \frac{z}{v_g}\,,\quad t_f = \frac{z}{v_f} \tag{3.2.4}$$

For example, $t_g = d\phi/d\omega = (d\beta/d\omega)z = z/v_g$, and similarly for the other ones, resulting in the definitions:

$$v_p = \frac{\omega}{\beta(\omega)}\,,\quad v_g = \frac{d\omega}{d\beta}\,,\quad v_f = \lim_{\omega\to\infty}\frac{\omega}{\beta(\omega)} \tag{3.2.5}$$

The expressions for the phase and group velocities agree with those of Sec. 1.11. Under the reasonable assumption that $\epsilon(\omega)\to\epsilon_0$ as $\omega\to\infty$, which is justified on the basis of the permittivity model of Eq. (1.9.13), we have $k(\omega) = \omega\sqrt{\epsilon(\omega)\mu_0} \to \omega\sqrt{\epsilon_0\mu_0} = \omega/c$, where c is the speed of light in vacuum. Therefore, the signal front velocity and front delay are:

$$v_f = \lim_{\omega\to\infty}\frac{\omega}{\beta(\omega)} = \lim_{\omega\to\infty}\frac{\omega}{\omega/c} = c \quad\Rightarrow\quad t_f = \frac{z}{c} \tag{3.2.6}$$

Thus, we expect that the impulse response $h(z,t)$ of the propagation medium would satisfy the causality condition:

$$h(z,t) = 0\,,\quad\text{for}\quad t < t_f = \frac{z}{c} \tag{3.2.7}$$

We show this below. More generally, if the input pulse at $z = 0$ vanishes for $t < t_0$, the propagated pulse to distance z will vanish for $t < t_0 + z/c$. This is the statement of *relativistic causality*, that is, if the input signal has a sharp, discontinuous, front at some time t_0, then that front cannot move faster than the speed of light in vacuum and cannot reach the point z faster than z/c seconds later. Mathematically,

$$E(0,t) = 0 \;\text{ for }\; t < t_0 \quad\Rightarrow\quad E(z,t) = 0 \;\text{ for }\; t < t_0 + \frac{z}{c} \tag{3.2.8}$$

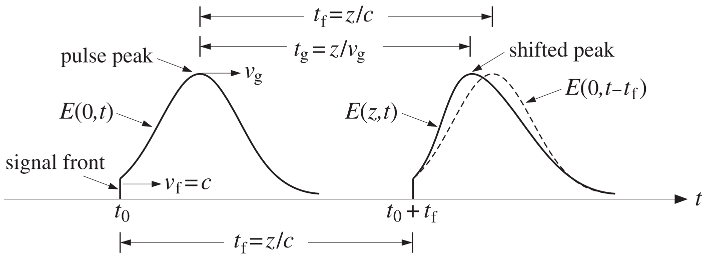


Fig. 3.2.1 Causal pulse propagation, but with superluminal group velocity $(v_g > c)$.

Fig. 3.2.1 depicts this property. Sommerfeld and Brillouin [154,1111] originally showed this property for a causal sinusoidal input, that is, $E(0,t)= e^{j\omega_0 t}u(t)$.

Group velocity describes the speed of the peak of the envelope of a signal and is a concept that applies only to narrow-band pulses. As mentioned in Sec. 1.11, it is possible that if this narrow frequency band is concentrated in the vicinity of an anomalous dispersion region, that is, near an absorption peak, the corresponding group velocity will exceed the speed of light in vacuum, $v_g > c$, or even become negative depending on the value of the negative slope of the refractive index $dn_r/d\omega < 0$.

Conventional wisdom has it that the condition $v_g > c$ is not at odds with relativity theory because the strong absorption near the resonance peak causes severe distortion and attenuation of the signal peak and the group velocity loses its meaning. However, in recent years it has been shown theoretically and experimentally [228,229,247] that the group velocity can retain its meaning as representing the speed of the peak even if v_g is superluminal or negative. Yet, relativistic causality is preserved because the signal front travels with the speed of light. It is the sharp discontinuous front of a signal that may convey information, not necessarily its peak. Because the pulse undergoes continuous reshaping as it propagates, the front cannot be overtaken by the faster moving peak.

This is explained pictorially in Fig. 3.2.1 which depicts such a case where $v_g > c$, and therefore, $t_g < t_f$. For comparison, the actual field $E(z,t)$ is shown together with the input pulse as if the latter had been traveling in vacuum, $E(0,t-z/c)$, reaching the point z with a delay of $t_f = z/c$. The peak of the pulse, traveling with speed v_g, gets delayed by the group delay t_g when it arrives at distance z. Because $t_g < t_f$, the peak of $E(z,t)$ shifts forward in time and occurs *earlier* than it would if the pulse were traveling in vacuum. Such peak shifting is a consequence of the "filtering" or "rephasing" taking place due to the propagator filter's frequency response $e^{-jk(\omega)z}$.

The causality conditions (3.2.7) and (3.2.8) imply that the value of the propagated field $E(z,t)$ at some time instant $t > t_0 + z/c$ is determined only by those values of the input pulse $E(0,t')$ that are z/c seconds earlier, that is, for $t_0 \le t' \le t - z/c$. This follows from the convolutional equation (3.1.7): the factor $h(z,t-t')$ requires that $t - t' \ge z/c$, the factor $E(0,t')$ requires $t' \ge t_0$, yielding $t_0 \le t' \le t - z/c$. Thus,

$$E(z,t)= \int_{t_0}^{t-z/c} h(z,t-t')E(0,t')dt'\,, \quad \text{for} \quad t > t_0 + z/c \tag{3.2.9}$$

For example, the value of $E(z,t)$ at $t = t_1 + t_f = t_1 + z/c$ is given by:

$$E(z,t_1+t_f)= \int_{t_0}^{t_1} h(z,t_1+t_f-t')E(0,t')dt'$$

Thus, as shown in Fig. 3.2.2, the shaded portion of the input $E(0,t')$ over the time interval $t_0 \le t' \le t_1$ determines causally the shaded portion of the propagated signal $E(z,t)$ over the interval $t_0 + t_f \le t \le t_1 + t_f$. The peaks, on the other hand, are not causally related. Indeed, the interval $[t_0, t_1]$ of the input does not include the peak, whereas the interval $[t_0 + t_f,\, t_1 + t_f]$ of the output does include the (shifted) peak.

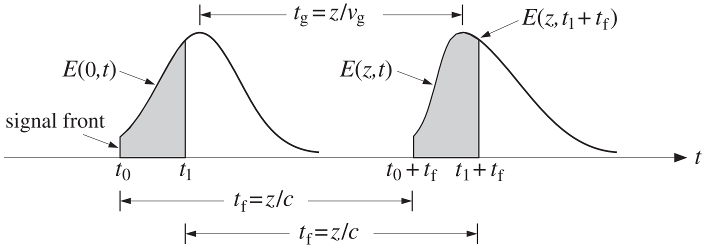


Fig. 3.2.2 Shaded areas show causally related portions of input and propagated signals.

Next, we provide a justification of Eq. (3.2.8). The condition $E(0,t)= 0$ for $t < t_0$, implies that its Fourier transform is:

$$\hat{E}(0,\omega)= \int_{t_0}^{\infty} e^{-j\omega t}E(0,t)dt \quad \Rightarrow \quad e^{j\omega t_0}\hat{E}(0,\omega)= \int_{0}^{\infty} e^{-j\omega t}E(0,t+t_0)dt \tag{3.2.10}$$

where the latter equation was obtained by the change of integration variable from t to $t+t_0$. It follows now that $e^{j\omega t_0}\hat{E}(0,\omega)$ is analytically continuable into the lower-half ω-plane. Indeed, the replacement $e^{-j\omega t}$ by $e^{-j(\omega - j\sigma)t} = e^{-\sigma t}e^{-j\omega t}$ with $\sigma > 0$ and $t > 0$, improves the convergence of the time integral in (3.2.10). We may write now Eq. (3.1.3) in the following form:

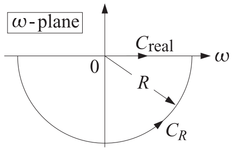
$$E(z,t)= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega t - \omega t_0 - kz)}\, e^{j\omega t_0}\hat{E}(0,\omega)d\omega \tag{3.2.11}$$

and assume that $t < t_0 + z/c$. A consequence of the permittivity model (1.9.13) is that the wavenumber $k(\omega)$ has singularities only in the upper-half ω-plane and is analytic in the lower half. For example, for the single-resonance case, we have:

$$\epsilon(\omega)= \epsilon_0\left[1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 + j\omega\gamma}\right] \quad \Rightarrow \quad \begin{aligned} \text{zeros} &= \frac{j\gamma}{2} \pm \sqrt{\omega_0^2 + \omega_p^2 - \frac{\gamma^2}{4}} \\ \text{poles} &= \frac{j\gamma}{2} \pm \sqrt{\omega_0^2 - \frac{\gamma^2}{4}} \end{aligned}$$

Thus, the integrand of Eq. (3.2.11) is analytic in the lower-half ω-plane and we may replace the integration path along the real axis by the lower semi-circular counterclockwise path C_R at a very large radius R, as shown below:

$$\begin{aligned} E(z,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j(\omega t-\omega t_0-kz)}\, e^{j\omega t_0}\hat{E}(0,\omega)\,d\omega \\ &= \lim_{R\to\infty}\frac{1}{2\pi}\int_{C_R} e^{j(\omega t-\omega t_0-kz)}\, e^{j\omega t_0}\hat{E}(0,\omega)\,d\omega \end{aligned}$$



But for large ω, we may replace $k(\omega)= \omega/c$. Thus,

$$E(z,t)= \lim_{R\to\infty}\frac{1}{2\pi}\int_{C_R} e^{j\omega(t-t_0-z/c)}\, e^{j\omega t_0}\hat{E}(0,\omega)\,d\omega$$

Because $t - t_0 - z/c < 0$, and under the mild assumption that $e^{j\omega t_0}\hat{E}(0,\omega)\to 0$ for $|\omega| = R \to \infty$ in the lower-half plane, it follows from the Jordan lemma that the above integral will be zero. Therefore, $E(z,t)= 0$ for $t < t_0 + z/c$.

As an example, consider the signal $E(0,t)= e^{-a(t-t_0)}e^{j\omega_0(t-t_0)}u(t-t_0)$, that is, a delayed exponentially decaying ($a > 0$) causal sinusoid. Its Fourier transform is

$$\hat{E}(0,\omega)= \frac{e^{-j\omega t_0}}{j(\omega-\omega_0-ja)} \quad\Rightarrow\quad e^{j\omega t_0}\hat{E}(0,\omega)= \frac{1}{j(\omega-\omega_0-ja)}$$

which is analytic in the lower half-plane and converges to zero for $|\omega| \to \infty$.

The proof of Eq. (3.2.7) is similar. Because of the analyticity of $k(\omega)$, the integration path in Eq. (3.1.6) can again be replaced by C_R, and $k(\omega)$ replaced by ω/c:

$$h(z,t)= \lim_{R\to\infty}\frac{1}{2\pi}\int_{C_R} e^{j\omega(t-z/c)}d\omega\,, \quad \text{for}\ \ t < z/c$$

This integral can be done exactly,[†] and leads to a standard representation of the delta function:

$$h(z,t)= \lim_{R\to\infty}\frac{\sin\big(R(t-z/c)\big)}{\pi(t-z/c)} = \delta(t-z/c)$$

which vanishes since we assumed that $t < z/c$. For $t > z/c$, the contour in (3.1.6) can be closed in the upper half-plane, but its evaluation requires knowledge of the particular singularities of $k(\omega)$.

3.3 Exact Impulse Response Examples

Some exactly solvable examples are given in [161]. They are all based on the following Fourier transform pair, which can be found in [156]:[‡]

$$\begin{aligned} H(z,\omega) &= e^{-jk(\omega)z} = e^{-t_f\sqrt{j\omega+a+b}\sqrt{j\omega+a-b}} \\ h(z,t) &= \delta(t-t_f)e^{-at_f} + \frac{I_1\Big(b\sqrt{t^2-t_f^2}\Big)}{\sqrt{t^2-t_f^2}}\,bt_f\,e^{-at}\,u(t-t_f) \end{aligned} \tag{3.3.1}$$

[†] set $\omega = Re^{j\theta}$, $d\omega = jRe^{j\theta}d\theta$, and integrate over $-\pi \le \theta \le 0$

[‡] see the pair 863.1 on p. 110 of [156].

where $I_1(x)$ is the modified Bessel function of the first kind of order one, and $t_f = z/c$ is the front delay. The unit step $u(t-t_f)$ enforces the causality condition (3.2.7). From the expression of $H(z,\omega)$, we identify the corresponding wavenumber:

$$k(\omega)= \frac{-j}{c}\sqrt{j\omega+a+b}\,\sqrt{j\omega+a-b} \tag{3.3.2}$$

The following physical examples are described by appropriate choices of the parameters a, b, c in Eq. (3.3.2):

1. $a = 0\,,\quad b = 0$ – propagation in vacuum or dielectric
2. $a > 0\,,\quad b = 0$ – weakly conducting dielectric
3. $a = b > 0$ – medium with finite conductivity
4. $a = 0\,,\quad b = j\omega_p$ – lossless plasma
5. $a = 0\,,\quad b = j\omega_c$ – hollow metallic waveguide
6. $a + b = R'/L'\,,\quad a - b = G'/C'$ – lossy transmission line

The anti-hermitian property $k(-\omega)^* = -k(\omega)$ is satisfied in two cases: when the parameters a, b are both real, or, when a is real and b imaginary.

In case 1, we have $k = \omega/c$ and $h(z,t)= \delta(t-t_f)= \delta(t-z/c)$. Setting $a = c\alpha > 0$ and $b = 0$, we find for case 2:

$$k = \frac{\omega - ja}{c} = \frac{\omega}{c} - j\alpha \tag{3.3.3}$$

which corresponds to a medium with a constant attenuation coefficient $\alpha = a/c$ and a propagation constant $\beta = \omega/c$, as was the case of a weakly conducting dielectric of Sec. 2.7. In this case c is the speed of light in the dielectric, i.e. $c = 1/\sqrt{\mu\epsilon}$ and a is related to the conductivity σ by $a = c\alpha = \sigma/2\epsilon$. The medium impulse response is:

$$h(z,t)= \delta(t-t_f)e^{-at_f} = \delta(t-z/c)e^{-\alpha z}$$

Eq. (3.1.7) then implies that an input signal will travel at speed c while attenuating with distance:

$$E(z,t)= e^{-\alpha z}E(0,t-z/c)$$

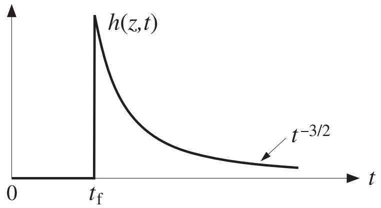
Case 3 describes a medium with frequency-independent permittivity and conductivity ϵ, σ with the parameters $a = b = \sigma/2\epsilon$ and $c = 1/\sqrt{\mu_0\epsilon}$. Eq. (3.3.2) becomes:

$$k = \frac{\omega}{c}\sqrt{1 - j\frac{\sigma}{\omega\epsilon}} \tag{3.3.4}$$

and the impulse response is:

$$h(z,t)= \delta(t-z/c)e^{-az/c} + \frac{I_1\Big(a\sqrt{t^2-(z/c)^2}\Big)}{\sqrt{t^2-(z/c)^2}}\,\frac{az}{c}\,e^{-at}\,u(t-z/c) \tag{3.3.5}$$

A plot of $h(z,t)$ for $t > t_f$ is shown below.



For large t, $h(z,t)$ is not exponentially decaying, but falls like $1/t^{3/2}$. Using the large-x asymptotic form $I_1(x)\to e^x/\sqrt{2\pi x}$, and setting $\sqrt{t^2-t_f^2}\to t$ for $t\gg t_f$, we find

$$h(z,t)\to \frac{e^{at}}{t\sqrt{2\pi at}}\,at_f\,e^{-at}=\frac{at_f}{t\sqrt{2\pi at}}\,,\quad t\gg t_f$$

Case 4 has parameters $a=0$ and $b=j\omega_p$ and describes propagation in a plasma, where ω_p is the plasma frequency. Eq. (3.3.2) reduces to Eq. (1.9.39):

$$k=\frac{1}{c}\sqrt{\omega^2-\omega_p^2}$$

To include evanescent waves (having $\omega<\omega_p$), Eq. (3.3.2) may be written in the more precise form that satisfies the required anti-hermitian property $k(-\omega)^*=-k(\omega)$:

$$k(\omega)=\frac{1}{c}\begin{cases}\operatorname{sign}(\omega)\sqrt{\omega^2-\omega_p^2}\,, & \text{if}\quad |\omega|\ge\omega_p\\ -j\sqrt{\omega_p^2-\omega^2}\,, & \text{if}\quad |\omega|\le\omega_p\end{cases}\tag{3.3.6}$$

When $|\omega|\le\omega_p$, the wave is evanescent in the sense that it attenuates exponentially with distance:

$$e^{-jkz}=e^{-z\sqrt{\omega_p^2-\omega^2}/c}$$

For numerical evaluation using MATLAB, it proves convenient to leave $k(\omega)$ in the form of Eq. (3.3.2), that is,

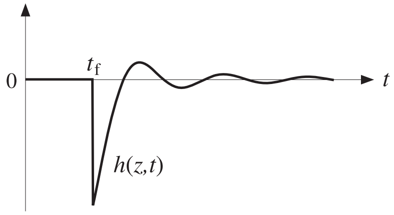
$$k(\omega)=\frac{-j}{c}\sqrt{j(\omega+\omega_p)}\,\sqrt{j(\omega-\omega_p)}$$

which evaluates correctly according to Eq. (3.3.6) using MATLAB's rules for computing square roots (e.g., $\sqrt{\pm j}=e^{\pm j\pi/4}$).

Because b is imaginary, we can use the property $I_1(jx)=jJ_1(x)$, where $J_1(x)$ is the ordinary Bessel function. Thus, setting $a=0$ and $b=j\omega_p$ in Eq. (3.3.1), we find:

$$h(z,t)=\delta(t-t_f)-\frac{J_1\Big(\omega_p\sqrt{t^2-t_f^2}\Big)}{\sqrt{t^2-t_f^2}}\,\omega_p t_f\,u(t-t_f)\tag{3.3.7}$$

A plot of $h(z,t)$ for $t>t_f$ is shown below.



The propagated output $E(z,t)$ due to a causal input, $E(0,t)=E(0,t)u(t)$, is obtained by convolution, where we must impose the conditions $t'\ge t_f$ and $t-t'\ge 0$:

$$E(z,t)=\int_{-\infty}^{\infty}h(z,t')E(0,t-t')\,dt'$$

which for $t\ge t_f$ leads to:

$$E(z,t)=E(0,t-t_f)-\int_{t_f}^{t}\frac{J_1\Big(\omega_p\sqrt{t'^2-t_f^2}\Big)}{\sqrt{t'^2-t_f^2}}\,\omega_p t_f\,E(0,t-t')\,dt'\tag{3.3.8}$$

We shall use Eq. (3.3.8) in the next section to illustrate the transient and steady-state response of a propagation medium such as a plasma or a waveguide. The large-t behavior of $h(z,t)$ is obtained from the asymptotic form:

$$J_1(x)\to\sqrt{\frac{2}{\pi x}}\,\cos\Big(x-\frac{3\pi}{4}\Big)\,,\quad x\gg 1$$

which leads to

$$h(z,t)\to-\frac{\sqrt{2\omega_p}\,t_f}{\sqrt{\pi}\,t^{3/2}}\,\cos\Big(\omega_p t-\frac{3\pi}{4}\Big)\,,\quad t\gg t_f\tag{3.3.9}$$

Case 5 is the same as case 4, but describes propagation in an air-filled hollow metallic waveguide with cutoff frequency ω_c. We will see in Chap. 9 that the dispersion relationship (3.3.6) is a consequence of the boundary conditions on the waveguide walls, and therefore, it is referred to as *waveguide dispersion*, as opposed to *material dispersion* arising from a frequency-dependent permittivity $\epsilon(\omega)$.

Case 6 describes a lossy transmission line (see Sec. 10.6) with distributed (that is, per unit length) inductance L', capacitance C', series resistance R', and shunt conductance G'. This case reduces to case 3 if $G'=0$. The corresponding propagation speed is $c=1/\sqrt{L'C'}$. The ω–k dispersion relationship can be written in the form of Eq. (10.6.5):

$$k=-j\sqrt{(R'+j\omega L')(G'+j\omega C')}=\omega\sqrt{L'C'}\sqrt{\Big(1-j\frac{R'}{\omega L'}\Big)\Big(1-j\frac{G'}{\omega C'}\Big)}$$

3.4 Transient and Steady-State Behavior

The frequency response $e^{-jk(\omega)z}$ is the Fourier transform of $h(z,t)$, but because of the causality condition $h(z,t)=0$ for $t<z/c$, the time-integration in this Fourier transform can be restricted to the interval $z/c<t<\infty$, that is,

$$\boxed{e^{-jk(\omega)z}=\int_{z/c}^{\infty}e^{-j\omega t}h(z,t)\,dt}\tag{3.4.1}$$

We mention, parenthetically, that Eq. (3.4.1), which incorporates the causality condition of $h(z,t)$, can be used to derive the lower half-plane analyticity of $k(\omega)$ and of the corresponding complex refractive index $n(\omega)$ defined through $k(\omega)=\omega n(\omega)/c$. The analyticity properties of $n(\omega)$ can then be used to derive the Kramers-Kronig dispersion relations satisfied by $n(\omega)$ itself [159], as opposed to those satisfied by the susceptibility $\chi(\omega)$ that were discussed in Sec. 1.10.

When a causal sinusoidal input is applied to the linear system $h(z,t)$, we expect the system to exhibit an initial transient behavior followed by the usual sinusoidal steady-state response. Indeed, applying the initial pulse $E(0,t)=e^{j\omega_0 t}u(t)$, we obtain from the system's convolutional equation:

$$E(z,t)=\int_{z/c}^{t} h(z,t')E(0,t-t')\,dt' = \int_{z/c}^{t} h(z,t')e^{j\omega_0(t-t')}\,dt'$$

where the restricted limits of integration follow from the conditions $t'\geq z/c$ and $t-t'\geq 0$ as required by the arguments of the functions $h(z,t')$ and $E(0,t-t')$. Thus, for $t\geq z/c$, the propagated field takes the form:

$$E(z,t)=e^{j\omega_0 t}\int_{z/c}^{t} e^{-j\omega_0 t'}h(z,t')\,dt' \tag{3.4.2}$$

In the steady-state limit, $t\to\infty$, the above integral tends to the frequency response (3.4.1) evaluated at $\omega=\omega_0$, resulting in the standard sinusoidal response:

$$e^{j\omega_0 t}\int_{z/c}^{t} e^{-j\omega_0 t'}h(z,t')\,dt' \to e^{j\omega_0 t}\int_{z/c}^{\infty} e^{-j\omega_0 t'}h(z,t')\,dt' = H(z,\omega_0)e^{j\omega_0 t}\,, \quad \text{or,}$$

$$E_{\text{steady}}(z,t)=e^{j\omega_0 t-jk(\omega_0)z}\,, \quad \text{for} \quad t\gg z/c \tag{3.4.3}$$

Thus, the field $E(z,t)$ eventually evolves into an ordinary plane wave at frequency ω_0 and wavenumber $k(\omega_0)=\beta(\omega_0)-j\alpha(\omega_0)$. The initial transients are given by the exact equation (3.4.2) and depend on the particular form of $k(\omega)$. They are generally referred to as "precursors" or "forerunners" and were originally studied by Sommerfeld and Brillouin [154,1111] for the case of a single-resonance Lorentz permittivity model.

It is beyond the scope of this book to study the precursors of the Lorentz model. However, we may use the exactly solvable model for a plasma or waveguide given in Eq. (3.3.7) and numerically integrate (3.4.2) to illustrate the transient and steady-state behavior.

Fig. 3.4.1 shows on the left graph the input sinusoid (dotted line) and the steady-state sinusoid (3.4.3) with k_0 computed from (3.3.6). The input and the steady output differ by the phase shift $-k_0 z$. The graph on the right shows the causal output for $t\geq t_f$ computed using Eq. (3.3.8) with the input $E(0,t)=\sin(\omega_0 t)u(t)$. During the initial transient period the output signal builds up to its steady-state form. The steady form of the left graph was not superimposed on the exact output because the two are virtually indistinguishable for large t. The graph units were arbitrary and we chose the following numerical values of the parameters:

$$c=1 \quad \omega_p=1\,, \quad \omega_0=3\,, \quad t_f=z=10$$

The following MATLAB code illustrates the computation of the exact and steady-state output signals:

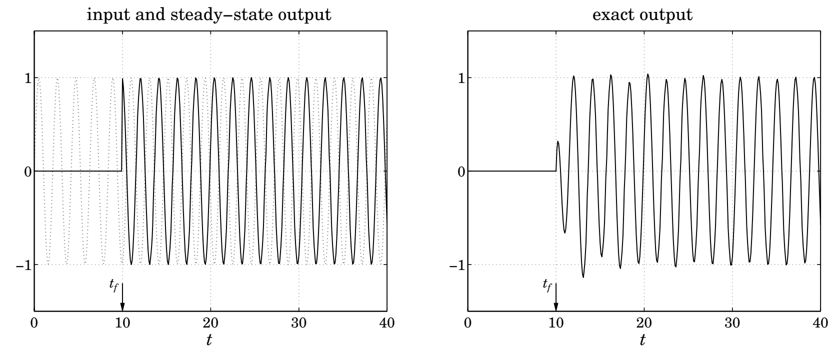


Fig. 3.4.1 Transient and steady-state sinusoidal response.

```
wp = 1; w0 = 3; tf = 10;
k0 = -j * sqrt(j*(w0+wp)) * sqrt(j*(w0-wp));      % equivalent to Eq. (3.3.6)

t = linspace(0,40, 401);

N = 15; K = 20;        % use N-point Gaussian quadrature, dividing [tf, t] into K subintervals

for i=1:length(t),
    if t(i)<tf,
       Ez(i) = 0;
       Es(i) = 0;
    else
       [w,x] = quadrs(linspace(tf,t(i),K), N);      % quadrature weights and points
       h = - wp^2 * tf * J1over(wp*sqrt(x.^2 - tf^2)) .* exp(j*w0*(t(i)-x));
       Ez(i) = exp(j*w0*(t(i)-tf)) + w'*h;          % exact output
       Es(i) = exp(j*w0*t(i)-j*k0*tf);              % steady-state
    end
end

es = imag(Es); ez = imag(Ez);                        % input is E(0,t) = sin(ω0 t) u(t)

figure; plot(t,es); figure; plot(t,ez);
```

The code uses the function `quadrs` (see Sec. 18.10 and Appendix I) to compute the integral over the interval $[t_f, t]$, dividing this interval into K subintervals and using an N-point Gauss-Legendre quadrature method on each subinterval.

We wrote a function `J1over` to implement the function $J_1(x)/x$. The function uses the power series expansion, $J_1(x)/x = 0.5(1-x^2/8+x^4/192)$, for small x, and the built-in MATLAB function `besselj` for larger x:

```
function y = J1over(x)

y = zeros(size(x));          % y has the same size as x

xmin = 1e-4;

i = find(abs(x) < xmin);
```

```
y(i) = 0.5 * (1 - x(i).^2 / 8 + x(i).^4 / 192);

i = find(abs(x) >= xmin);
y(i) = besselj(1, x(i)) ./ x(i);
```

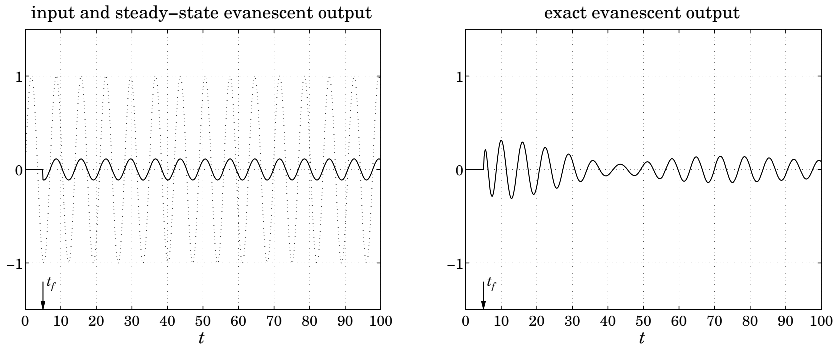


Fig. 3.4.2 Transient and steady-state response for evanescent sinusoids.

Fig. 3.4.2 illustrates an evanescent wave with $\omega_0 < \omega_p$. In this case the wavenumber becomes pure imaginary, $k_0 = -j\alpha_0 = -j\sqrt{\omega_p^2 - \omega_0^2}/c$, leading to an attenuated steady-state waveform:

$$E_{\text{steady}}(z,t) = e^{j\omega_0 t - jk_0 z} = e^{j\omega_0 t}e^{-\alpha_0 z}\,, \quad t \gg \frac{z}{c}$$

The following numerical values were used:

$$c = 1 \quad \omega_p = 1\,, \quad \omega_0 = 0.9\,, \quad t_f = z = 5$$

resulting in the imaginary wavenumber and attenuation amplitude:

$$k_0 = -j\alpha_0 = -0.4359j\,, \quad H_0 = e^{-jk_0 z} = e^{-\alpha_o z} = 0.1131$$

We chose a smaller value of z in order to get a reasonable value for the attenuated signal for display purposes. The left graph in Fig. 3.4.2 shows the input and the steady-state output signals. The right graph shows the exact output computed by the same MATLAB code given above. Again, we note that for large t (here, $t > 80$), the exact output approaches the steady one.

Finally, in Fig. 3.4.3 we illustrate the input-on and input-off transients for an input rectangular pulse of duration t_d, and for a causal gaussian pulse, that is,

$$E(0,t) = \sin(\omega_0 t)\left[u(t) - u(t - t_d)\right], \quad E(0,t) = e^{j\omega_0 t}\exp\left[-\frac{(t-t_c)^2}{2\tau_0^2}\right]u(t)$$

The input-off transients for the rectangular pulse are due to the oscillating and decaying tail of the impulse response $h(z,t)$ given in (3.3.9). The following values of the parameters were used:

$$c = 1 \quad \omega_p = 1\,, \quad \omega_0 = 3\,, \quad t_f = z = 30\,, \quad t_d = 20\,, \quad t_c = \tau_0 = 5$$

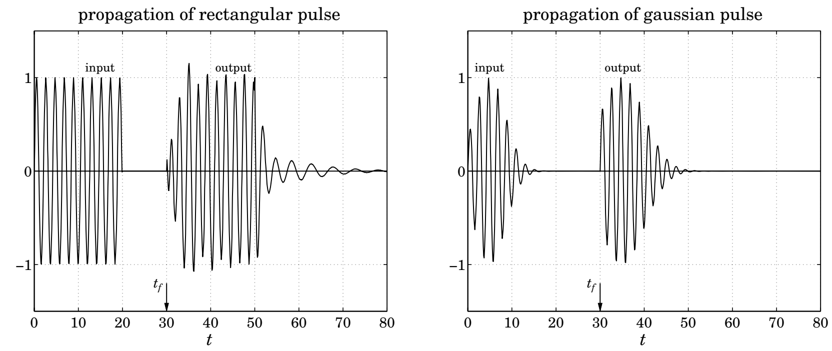


Fig. 3.4.3 Rectangular and gaussian pulse propagation.

The MATLAB code for the rectangular pulse case is essentially the same as above except that it uses the function `upulse` to enforce the finite pulse duration:

```
wp = 1; w0 = 3; tf = 30; td = 20; N = 15; K = 20;
k0 = -j * sqrt(j*(w0+wp)) * sqrt(j*(w0-wp));

t = linspace(0,80,801);

E0 = exp(j*w0*t) .* upulse(t,td);

for i=1:length(t),
  if t(i)<tf,
    Ez(i) = 0;
  else
    [w,x] = quadrs(linspace(tf,t(i),K), N);
    h = - wp^2 * tf * J1over(wp*sqrt(x.^2-tf^2)) .* ...
             exp(j*w0*(t(i)-x)) .* upulse(t(i)-x,td);
    Ez(i) = exp(j*w0*(t(i)-tf)).*upulse(t(i)-tf,td) + w'*h;
  end
end

e0 = imag(E0); ez = imag(Ez);

plot(t,ez,'-', t,e0,'-');
```

3.5 Pulse Propagation and Group Velocity

In this section, we show that the peak of a pulse travels with the group velocity. The concept of group velocity is associated with narrow-band pulses whose spectrum $\hat{E}(0,\omega)$ is narrowly concentrated in the neighborhood of some frequency, say, ω_0, with an effective frequency band $|\omega - \omega_0| \le \Delta\omega$, where $\Delta\omega \ll \omega_0$, as depicted in Fig. 3.5.1.

Such spectrum can be made up by translating a low-frequency spectrum, say $\hat{F}(0,\omega)$, to ω_0, that is, $\hat{E}(0,\omega) = \hat{F}(0,\omega - \omega_0)$. From the modulation property of Fourier trans-

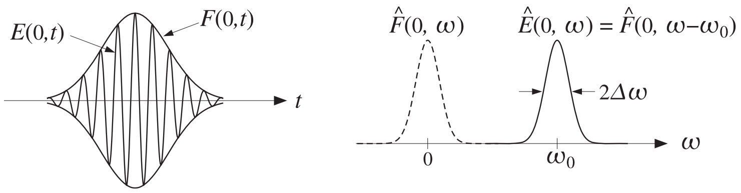


Fig. 3.5.1 High-frequency sinusoid with slowly-varying envelope.

forms, it follows that the corresponding time-domain signal $E(0,t)$ will be:

$$\hat{E}(0,\omega)=\hat{F}(0,\omega-\omega_0)\quad\Rightarrow\quad\boxed{E(0,t)=e^{j\omega_0 t}F(0,t)}\tag{3.5.1}$$

that is, a sinusoidal carrier modulated by a slowly varying envelope $F(0,t)$, where

$$F(0,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega' t}\hat{F}(0,\omega')\,d\omega'=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j(\omega-\omega_0)t}\hat{F}(0,\omega-\omega_0)\,d\omega\tag{3.5.2}$$

Because the integral over $\omega'=\omega-\omega_0$ is effectively restricted over the low-frequency band $|\omega'|\leq\Delta\omega$, the resulting envelope $F(0,t)$ will be slowly-varying (relative to the period $2\pi/\omega_0$ of the carrier.) If this pulse is launched into a dispersive medium with wavenumber $k(\omega)$, the propagated pulse to distance z will be given by:

$$E(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j(\omega t-kz)}\hat{F}(0,\omega-\omega_0)\,d\omega\tag{3.5.3}$$

Defining $k_0=k(\omega_0)$, we may rewrite $E(z,t)$ in the form of a modulated plane wave:

$$\boxed{E(z,t)=e^{j(\omega_0 t-k_0 z)}F(z,t)}\tag{3.5.4}$$

where the propagated envelope $F(z,t)$ is given by

$$\boxed{F(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j(\omega-\omega_0)t-j(k-k_0)z}\,\hat{F}(0,\omega-\omega_0)\,d\omega}\tag{3.5.5}$$

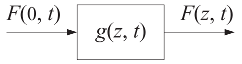
This can also be written in a convolutional form by defining the *envelope impulse response* function $g(z,t)$ in terms of the propagator impulse response $h(z,t)$:

$$h(z,t)=e^{j(\omega_0 t-k_0 z)}g(z,t)\tag{3.5.6}$$

so that

$$g(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j(\omega-\omega_0)t-j(k-k_0)z}\,d\omega\tag{3.5.7}$$

Then, the propagated envelope can be obtained by the convolutional operation:



$$F(z,t)=\int_{-\infty}^{\infty}g(z,t')F(0,t-t')\,dt'\tag{3.5.8}$$

Because $\hat{F}(0,\omega-\omega_0)$ restricts the effective range of integration in Eq. (3.5.5) to a narrow band about ω_0, one can expand $k(\omega)$ to a Taylor series about ω_0 and keep only the first few terms:

$$\boxed{k(\omega)=k_0+k_0'(\omega-\omega_0)+\frac{1}{2}k_0''(\omega-\omega_0)^2+\cdots}\tag{3.5.9}$$

where

$$k_0=k(\omega_0)\,,\quad k_0'=\frac{dk}{d\omega}\bigg|_{\omega_0}\,,\quad k_0''=\frac{d^2k}{d\omega^2}\bigg|_{\omega_0}\tag{3.5.10}$$

If $k(\omega)$ is real, we recognize k_0' as the inverse of the group velocity at frequency ω_0:

$$k_0'=\frac{dk}{d\omega}\bigg|_{\omega_0}=\frac{1}{v_g}\tag{3.5.11}$$

If k_0' is complex-valued, $k_0'=\beta_0'-j\alpha_0'$, then its real part determines the group velocity through $\beta_0'=1/v_g$, or, $v_g=1/\beta_0'$. The second derivative k_0'' is referred to as the "dispersion coefficient" and is responsible for the spreading and chirping of the wave packet, as we see below.

Keeping up to the quadratic term in the quantity $k(\omega)-k_0$ in (3.5.5), and changing integration variables to $\omega'=\omega-\omega_0$, we obtain the approximation:

$$F(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega'(t-k_0'z)-jk_0''z\omega'^2/2}\hat{F}(0,\omega')\,d\omega'\tag{3.5.12}$$

In the *linear* approximation, we may keep k_0' and ignore the k_0'' term, and in the *quadratic* approximation, we keep both k_0' and k_0''. For the linear case, we have by comparing with Eq. (3.5.2):

$$F(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega'(t-k_0'z)}\hat{F}(0,\omega')\,d\omega'=F(0,t-k_0'z)\tag{3.5.13}$$

Thus, assuming that k_0' is real so that $k_0'=1/v_g$, Eq. (3.5.13) implies that the initial envelope $F(0,t)$ is moving as whole with the group velocity v_g. The field $E(z,t)$ is obtained by modulating the high-frequency plane wave $e^{j(\omega_0 t-k_0 z)}$ with this envelope:

$$\boxed{E(z,t)=e^{j(\omega_0 t-k_0 z)}\,F(0,t-z/v_g)}\tag{3.5.14}$$

Every point on the envelope travels at the same speed v_g, that is, its shape remains unchanged as it propagates, as shown in Fig. 3.5.2. The high-frequency carrier suffers a phase-shift given by $-k_0 z$.

Similar approximations can be introduced in (3.5.7) anticipating that (3.5.8) will be applied only to narrowband input envelope signals $F(0,t)$:

$$g(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega'(t-k_0'z)-jk_0''z\omega'^2/2}\,d\omega'\tag{3.5.15}$$

This integral can be done exactly, and leads to the following expressions in the linear and quadratic approximation cases (assuming that k_0', k_0'' are real):

$$\begin{aligned}&\text{linear:} && g(z,t)=\delta(t-k_0'z)\\ &\text{quadratic:} && g(z,t)=\frac{1}{\sqrt{2\pi jk_0''z}}\exp\left[-\frac{(t-k_0'z)^2}{2jk_0''z}\right]\end{aligned}\tag{3.5.16}$$

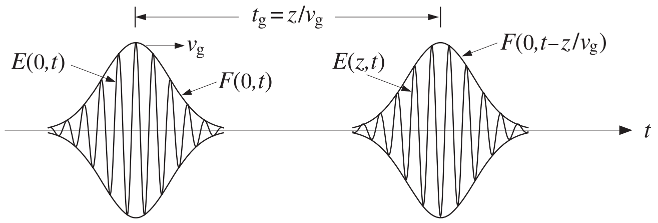


Fig. 3.5.2 Pulse envelope propagates with velocity v_g remaining unchanged in shape.

The corresponding frequency responses follow from Eq. (3.5.15), replacing ω' by ω:

$$\begin{aligned} &\text{linear:} && G(z,\omega)=e^{-jk_0' z\omega} \\ &\text{quadratic:} && G(z,\omega)=e^{-jk_0' z\omega}e^{-jk_0'' z\omega^2/2} \end{aligned} \tag{3.5.17}$$

The linear case is obtained from the quadratic one in the limit $k_0''\to 0$. We note that the integral of Eq. (3.5.15), as well as the gaussian pulse examples that we consider later, are special cases of the following Fourier integral:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t-(a+jb)\omega^2/2}\,d\omega = \frac{1}{\sqrt{2\pi(a+jb)}}\exp\left[-\frac{t^2}{2(a+jb)}\right] \tag{3.5.18}$$

where a, b are real, with the restriction that $a\geq 0$.[†] The integral for $g(z,t)$ corresponds to the case $a=0$ and $b=k_0'' z$. Using (3.5.16) into (3.5.8), we obtain Eq. (3.5.13) in the linear case and the following convolutional expression in the quadratic one:

$$\begin{aligned} &\text{linear:} && F(z,t)=F(0,t-k_0' z) \\ &\text{quadratic:} && F(z,t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi jk_0'' z}}\exp\left[-\frac{(t'-k_0' z)^2}{2jk_0'' z}\right]F(0,t-t')\,dt' \end{aligned} \tag{3.5.19}$$

and in the frequency domain:

$$\begin{aligned} &\text{linear:} && \hat{F}(z,\omega)=G(z,\omega)\hat{F}(0,\omega)=e^{-jk_0' z\omega}\hat{F}(0,\omega) \\ &\text{quadratic:} && \hat{F}(z,\omega)=G(z,\omega)\hat{F}(0,\omega)=e^{-jk_0' z\omega-jk_0'' z\omega^2/2}\hat{F}(0,\omega) \end{aligned} \tag{3.5.20}$$

3.6 Group Velocity Dispersion and Pulse Spreading

In the linear approximation, the envelope propagates with the group velocity v_g, remaining unchanged in shape. But in the quadratic approximation, as a consequence of Eq. (3.5.19), it spreads and reduces in amplitude with distance z, and it chirps. To see this, consider a gaussian input pulse of effective width τ_0:

$$F(0,t)=\exp\left[-\frac{t^2}{2\tau_0^2}\right] \quad\Rightarrow\quad E(0,t)=e^{j\omega_0 t}F(0,t)=e^{j\omega_0 t}\exp\left[-\frac{t^2}{2\tau_0^2}\right] \tag{3.6.1}$$

[†] Given the polar form $a+jb=Re^{j\theta}$, we must choose the square root $\sqrt{a+jb}=R^{1/2}e^{j\theta/2}$.

with Fourier transforms $\hat{F}(0,\omega)$ and $\hat{E}(0,\omega)=\hat{F}(0,\omega-\omega_0)$:

$$\hat{F}(0,\omega)=\sqrt{2\pi\tau_0^2}\,e^{-\tau_0^2\omega^2/2} \quad\Rightarrow\quad \hat{E}(0,\omega)=\sqrt{2\pi\tau_0^2}\,e^{-\tau_0^2(\omega-\omega_0)^2/2} \tag{3.6.2}$$

with an effective width $\Delta\omega = 1/\tau_0$. Thus, the condition $\Delta\omega \ll \omega_0$ requires that $\tau_0\omega_0 \gg 1$, that is, an envelope with a long duration relative to the carrier's period.

The propagated envelope $F(z,t)$ can be determined either from Eq. (3.5.19) or from (3.5.20). Using the latter, we have:

$$\hat{F}(z,\omega)=\sqrt{2\pi\tau_0^2}\,e^{-jk_0' z\omega-jk_0'' z\omega^2/2}e^{-\tau_0^2\omega^2/2}=\sqrt{2\pi\tau_0^2}\,e^{-jk_0' z\omega}e^{-(\tau_0^2+jk_0'' z)\omega^2/2} \tag{3.6.3}$$

The Fourier integral (3.5.18), then, gives the propagated envelope in the time domain:

$$F(z,t)=\sqrt{\frac{\tau_0^2}{\tau_0^2+jk_0'' z}}\,\exp\left[-\frac{(t-k_0' z)^2}{2(\tau_0^2+jk_0'' z)}\right] \tag{3.6.4}$$

Thus, effectively we have the replacement $\tau_0^2\to\tau_0^2+jk_0'' z$. Assuming for the moment that k_0' and k_0'' are real, we find for the magnitude of the propagated pulse:

$$|F(z,t)| = \left[\frac{\tau_0^4}{\tau_0^4+(k_0'' z)^2}\right]^{1/4}\exp\left[-\frac{(t-k_0' z)^2\,\tau_0^2}{2\big(\tau_0^4+(k_0'' z)^2\big)}\right] \tag{3.6.5}$$

where we used the property $|\tau_0^2+jk_0'' z| = \sqrt{\tau_0^4+(k_0'' z)^2}$. The effective width is determined from the argument of the exponent to be:

$$\tau^2 = \frac{\tau_0^4+(k_0'' z)^2}{\tau_0^2} \quad\Rightarrow\quad \tau = \left[\tau_0^2+\left(\frac{k_0'' z}{\tau_0}\right)^2\right]^{1/2} \tag{3.6.6}$$

Therefore, the pulse width increases with distance z. Also, the amplitude of the pulse decreases with distance, as measured for example at the peak maximum:

$$|F|_{\max} = \left[\frac{\tau_0^4}{\tau_0^4+(k_0'' z)^2}\right]^{1/4}$$

The peak maximum occurs at the group delay $t=k_0' z$, and hence it is moving at the group velocity $v_g=1/k_0'$.

The effect of pulse spreading and amplitude reduction due to the term k_0'' is referred to as *group velocity dispersion* or *chromatic dispersion*. Fig. 3.6.1 shows the amplitude decrease and spreading of the pulse with distance, as well as the chirping effect (to be discussed in the next section.)

Because the frequency width is $\Delta\omega = 1/\tau_0$, we may write the excess time spread $\Delta\tau = k_0'' z/\tau_0$ in the form $\Delta\tau = k_0'' z\Delta\omega$. This can be understood in terms of the change in the group delay. It follows from $t_g = z/v_g = k'z$ that the change in t_g due to $\Delta\omega$ will be:

$$\Delta t_g = \frac{dt_g}{d\omega}\,\Delta\omega = \frac{dk'}{d\omega}\,z\,\Delta\omega = \frac{d^2k}{d\omega^2}\,z\,\Delta\omega = k''z\,\Delta\omega \tag{3.6.7}$$

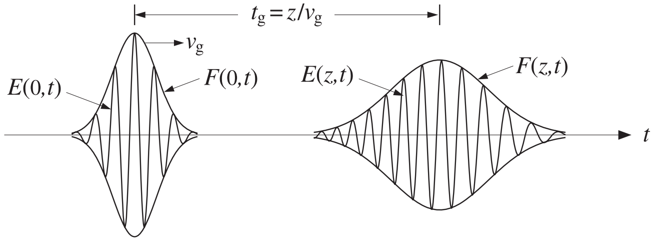


Fig. 3.6.1 Pulse spreading and chirping.

which can also be expressed in terms of the free-space wavelength $\lambda = 2\pi c/\omega$:

$$\Delta t_g = \frac{dt_g}{d\lambda}\,\Delta\lambda = \frac{dk'}{d\lambda}\,z\,\Delta\lambda = D\,z\,\Delta\lambda \tag{3.6.8}$$

where D is the "dispersion coefficient"

$$D = \frac{dk'}{d\lambda} = -\frac{2\pi c}{\lambda^2}\,\frac{dk'}{d\omega} = -\frac{2\pi c}{\lambda^2}\,k'' \tag{3.6.9}$$

where we replaced $d\lambda = -(\lambda^2/2\pi c)d\omega$. Since k' is related to the group refractive index n_g by $k' = 1/v_g = n_g/c$, we may obtain an alternative expression for D directly in terms of the refractive index n. Using Eq. (1.11.6), that is, $n_g = n - \lambda dn/d\lambda$, we find

$$D = \frac{dk'}{d\lambda} = \frac{1}{c}\,\frac{dn_g}{d\lambda} = \frac{1}{c}\,\frac{d}{d\lambda}\left[n - \lambda\frac{dn}{d\lambda}\right] = -\frac{\lambda}{c}\,\frac{d^2n}{d\lambda^2} \tag{3.6.10}$$

Combining Eqs. (3.6.9) and (3.6.10), we also have:

$$k'' = \frac{\lambda^3}{2\pi c^2}\,\frac{d^2n}{d\lambda^2} \tag{3.6.11}$$

In digital data transmission using optical fibers, the issue of pulse broadening as measured by (3.6.8) becomes important because it limits the maximum usable bit rate, or equivalently, the maximum propagation distance. The interpulse time interval of, say, T_b seconds by which bit pulses are separated corresponds to a data rate of $f_b = 1/T_b$ bits/second and must be longer than the broadening time, $T_b > \Delta t_g$, otherwise the broadened pulses will begin to overlap preventing their clear identification as separate. This limits the propagation distance z to a maximum value:[†]

$$D\,z\,\Delta\lambda \le T_b = \frac{1}{f_b} \quad\Rightarrow\quad \boxed{z \le \frac{1}{f_b\,D\,\Delta\lambda} = \frac{1}{f_b\,k''\Delta\omega}} \tag{3.6.12}$$

Because $D = \Delta t_g/z\Delta\lambda$, the parameter D is typically measured in units of picoseconds per km per nanometer—the km referring to the distance z and the nm to the wavelength spread $\Delta\lambda$. Similarly, the parameter $k'' = \Delta t_g/z\Delta\omega$ is measured in units of ps^2/km. As an example, we used the Sellmeier model for fused silica given in Eq. (1.9.18)

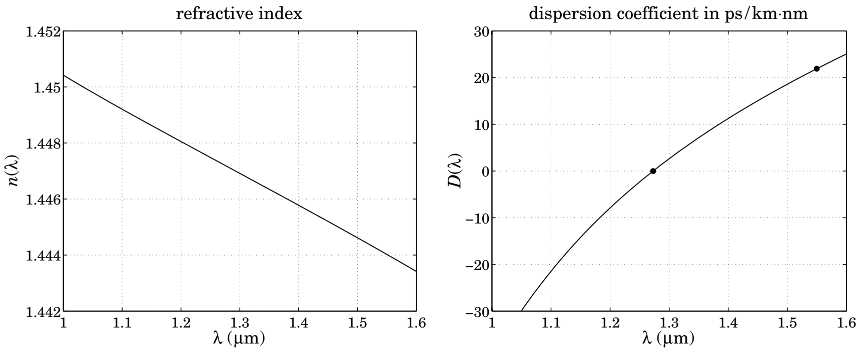


Fig. 3.6.2 Refractive index and dispersion coefficient of fused silica.

to plot in Fig. 3.6.2 the refractive index $n(\lambda)$ and the dispersion coefficient $D(\lambda)$ versus wavelength in the range $1 \le \lambda \le 1.6$ μm.

We observe that D vanishes, and hence also $k'' = 0$, at about $\lambda = 1.27$ μm corresponding to dispersionless propagation. This wavelength is referred to as a "zero dispersion wavelength." However, the preferred wavelength of operation is $\lambda = 1.55$ μm at which fiber losses are minimized. At $\lambda = 1.55$, we calculate the following refractive index values from the Sellmeier equation:

$$n = 1.444\,,\quad \frac{dn}{d\lambda} = -11.98{\times}10^{-3}\ \mu\text{m}^{-1}\,,\quad \frac{d^2n}{d\lambda^2} = -4.24{\times}10^{-3}\ \mu\text{m}^{-2} \tag{3.6.13}$$

resulting in the group index $n_g = 1.463$ and group velocity $v_g = c/n_g = 0.684c$. Using (3.6.10) and (3.6.11), the calculated values of D and k'' are:

$$D = 21.9\ \frac{\text{ps}}{\text{km}\cdot\text{nm}}\,,\quad k'' = -27.9\ \frac{\text{ps}^2}{\text{km}} \tag{3.6.14}$$

The ITU-G.652 standard single-mode fiber [206] has the following nominal values of the dispersion parameters at $\lambda = 1.55$ μm:

$$D = 17\ \frac{\text{ps}}{\text{km}\cdot\text{nm}}\,,\quad k'' = -21.67\ \frac{\text{ps}^2}{\text{km}} \tag{3.6.15}$$

with the dispersion coefficient $D(\lambda)$ given approximately by the fitted linearized form in the neighborhood of 1.55 μm:

$$D(\lambda) = 17 + 0.056(\lambda - 1550)\ \frac{\text{ps}}{\text{km}\cdot\text{nm}}\,,\quad \text{with } \lambda \text{ in units of nm}$$

Moreover, the standard fiber has a zero-dispersion wavelength of about 1.31 μm and an attenuation constant of about 0.2 dB/km.

We can use the values in (3.6.15) to get a rough estimate of the maximum propagation distance in a standard fiber. We assume that the data rate is $f_b = 40$ Gbit/s, so that the interpulse spacing is $T_b = 25$ ps. For a 10 picosecond pulse, i.e., $\tau_0 = 10$ ps and $\Delta\omega = 1/\tau_0 = 0.1$ rad/ps, we estimate the wavelength spread to be $\Delta\lambda = (\lambda^2/2\pi c)\Delta\omega = 0.1275$ nm at $\lambda = 1.55\ \mu\text{m}$. Using Eq. (3.6.12), we find the limit $z \le 11.53$ km—a distance that falls short of the 40-km and 80-km recommended lengths.

[†] where the absolute values of D, k'' must be used in Eq. (3.6.12).

Longer propagation lengths can be achieved by using dispersion compensation techniques, such as using chirped inputs or adding negative-dispersion fiber lengths. We discuss chirping and dispersion compensation in the next two sections.

The result (3.6.4) remains valid [163], with some caveats, when the wavenumber is complex valued, $k(\omega) = \beta(\omega) - j\alpha(\omega)$. The parameters $k_0' = \beta_0' - j\alpha_0'$ and $k_0'' = \beta_0'' - j\alpha_0''$ can be substituted in Eqs. (3.6.3) and (3.6.4):

$$\begin{gathered}\hat{F}(z,\omega) = \sqrt{2\pi\tau_0^2}\, e^{-j(\beta_0'-j\alpha_0')z\omega}\, e^{-\left(\tau_0^2+(\alpha_0''+j\beta_0'')z\right)\omega^2/2} \\ F(z,t) = \sqrt{\frac{\tau_0^2}{\tau_0^2+\alpha_0''z+j\beta_0''z}}\,\exp\left[-\frac{\big(t-(\beta_0'-j\alpha_0')z\big)^2}{2(\tau_0^2+\alpha_0''z+j\beta_0''z)}\right]\end{gathered} \tag{3.6.16}$$

The Fourier integral (3.5.18) requires that the real part of the effective complex width $\tau_0^2 + jk_0''z = (\tau_0^2 + \alpha_0''z) + j\beta_0''z$ be positive, that is, $\tau_0^2 + \alpha_0''z > 0$. If α_0'' is negative, this condition limits the distances z over which the above approximations are valid. The exponent can be written in the form:

$$-\frac{\big(t-(\beta_0'-j\alpha_0')z\big)^2}{2(\tau_0^2+\alpha_0''z+j\beta_0''z)} = -\frac{(t-\beta_0'z+j\alpha_o'z)^2(\tau_0^2+\alpha_0''z-j\beta_0''z)}{2\big[(\tau_0^2+\alpha_0''z)^2+(\beta_0''z)^2\big]} \tag{3.6.17}$$

Separating this into its real and imaginary parts, one can show after some algebra that the magnitude of $F(z,t)$ is given by:[†]

$$|F(z,t)| = \left[\frac{\tau_0^4}{(\tau_0^2+\alpha_0''z)^2+(\beta_0''z)^2}\right]^{1/4} \exp\left[\frac{\alpha_0'^2z^2}{2(\tau_0^2+\alpha_0''z)}\right]\cdot\exp\left[-\frac{(t-t_g)^2}{2\tau^2}\right] \tag{3.6.18}$$

where the peak of the pulse does not quite occur at the ordinary group delay $t_g = \beta_0'z$, but rather at the effective group delay:

$$t_g = \beta_0'z - \frac{\alpha_0'\beta_0''z^2}{\tau_0^2+\alpha_0''z}$$

The effective width of the peak generalizes Eq. (3.6.6)

$$\tau^2 = \tau_0^2 + \alpha_0''z + \frac{(\beta_0''z)^2}{\tau_0^2+\alpha_0''z}$$

From the imaginary part of Eq. (3.6.17), we observe two additional effects. First, the non-zero coefficient of the jt term is equivalent to a z-dependent frequency shift of the carrier frequency ω_0, and second, from the coefficient of $jt^2/2$, there will be a certain amount of chirping as discussed in the next section. The frequency shift and chirping coefficient (generalizing Eq. (3.7.6)) turn out to be:

$$\Delta\omega_0 = -\frac{\alpha_o'z(\tau_0^2+\alpha_0''z)}{(\tau_0^2+\alpha_0''z)^2+(\beta_0''z)^2}\,,\qquad \dot{\omega}_0 = \frac{\beta_0''z}{(\tau_0^2+\alpha_0''z)^2+(\beta_0''z)^2}$$

[†] note that if $F = Ae^B$, then $|F| = |A|e^{\mathrm{Re}(B)}$.

In most applications and in the fast and slow light experiments that have been carried out thus far, care has been taken to minimize these effects by operating in frequency bands where α_0', α_0'' are small and by limiting the propagation distance z.

3.7 *Propagation and Chirping*

A chirped sinusoid has an instantaneous frequency that changes *linearly* with time, referred to as linear frequency modulation (FM). It is obtained by the substitution:

$$e^{j\omega_0 t} \quad\to\quad e^{j(\omega_0 t + \dot{\omega}_0 t^2/2)} \tag{3.7.1}$$

where the "chirping parameter" $\dot{\omega}_0$ is a constant representing the rate of change of the instantaneous frequency. The phase $\theta(t)$ and instantaneous frequency $\dot{\theta}(t) = d\theta(t)/dt$ are for the above sinusoids:

$$\begin{aligned}\theta(t) &= \omega_0 t &&\to\quad \theta(t) = \omega_0 t + \frac{1}{2}\dot{\omega}_0 t^2 \\ \dot{\theta}(t) &= \omega_0 &&\to\quad \dot{\theta}(t) = \omega_0 + \dot{\omega}_0 t\end{aligned} \tag{3.7.2}$$

The parameter $\dot{\omega}_0$ can be positive or negative resulting in an increasing or decreasing instantaneous frequency. A chirped gaussian pulse is obtained by modulating a chirped sinusoid by a gaussian envelope:

$$E(0,t) = e^{j(\omega_0 t+\dot{\omega}_0 t^2/2)}\exp\left[-\frac{t^2}{2\tau_0^2}\right] = e^{j\omega_0 t}\exp\left[-\frac{t^2}{2\tau_0^2}(1-j\dot{\omega}_0\tau_0^2)\right] \tag{3.7.3}$$

which can be written in the following form, in the time and frequency domains:

$$E(0,t) = e^{j\omega_0 t}\exp\left[-\frac{t^2}{2\tau_{\text{chirp}}^2}\right] \quad\Leftrightarrow\quad \hat{E}(0,\omega) = \sqrt{2\pi\tau_{\text{chirp}}^2}\,e^{-\tau_{\text{chirp}}^2(\omega-\omega_0)^2/2} \tag{3.7.4}$$

where τ_{chirp}^2 is an equivalent complex-valued width parameter defined by:

$$\tau_{\text{chirp}}^2 = \frac{\tau_0^2}{1-j\dot{\omega}_0\tau_0^2} = \frac{\tau_0^2(1+j\dot{\omega}_0\tau_0^2)}{1+\dot{\omega}_0^2\tau_0^4} \tag{3.7.5}$$

Thus, a complex-valued width is associated with linear chirping. An unchirped gaussian pulse that propagates by a distance z into a medium becomes chirped because it acquires a complex-valued width, that is, $\tau_0^2 + jk_0''z$, as given by Eq. (3.6.4). Therefore, propagation is associated with chirping. Close inspection of Fig. 3.6.1 reveals that the front of the pulse appears to have a higher carrier frequency than its back (in this figure, we took $k_0'' < 0$, for normal dispersion). The effective chirping parameter $\dot{\omega}_0$ can be identified by writing the propagated envelope in the form:

$$\begin{aligned}F(z,t) &= \sqrt{\frac{\tau_0^2}{\tau_0^2+jk_0''z}}\,\exp\left[-\frac{(t-k_0'z)^2}{2(\tau_0^2+jk_0''z)}\right] \\ &= \sqrt{\frac{\tau_0^2}{\tau_0^2+jk_0''z}}\,\exp\left[-\frac{(t-k_0'z)^2}{2\big(\tau_0^4+(k_0''z)^2\big)}(\tau_0^2-jk_0''z)\right]\end{aligned}$$

Comparing with (3.7.3), we identify the chirping parameter due to propagation:

$$\dot{\omega}_0 = \frac{k_0''z}{\tau_0^4 + (k_0''z)^2} \tag{3.7.6}$$

If a *chirped gaussian input* is launched into a propagation medium, then the chirping due to propagation will combine with the input chirping. The two effects can sometimes cancel each other leading to *pulse compression* rather than spreading. Indeed, if the chirped pulse (3.7.4) is propagated by a distance z, then according to (3.6.4), the propagated envelope will be:

$$F(z,t) = \sqrt{\frac{\tau_{\text{chirp}}^2}{\tau_{\text{chirp}}^2 + jk_0''z}}\,\exp\left[-\frac{(t-k_0'z)^2}{2(\tau_{\text{chirp}}^2 + jk_0''z)}\right] \tag{3.7.7}$$

The effective complex-valued width parameter will be:

$$\tau_{\text{chirp}}^2 + jk_0''z = \frac{\tau_0^2(1+j\dot{\omega}_0\tau_0^2)}{1+\dot{\omega}_0^2\tau_0^4} + jk_0''z = \frac{\tau_0^2}{1+\dot{\omega}_0^2\tau_0^4} + j\left(\frac{\dot{\omega}_0\tau_0^4}{1+\dot{\omega}_0^2\tau_0^4} + k_0''z\right) \tag{3.7.8}$$

If $\dot{\omega}_0$ is selected such that

$$\frac{\dot{\omega}_0\tau_0^4}{1+\dot{\omega}_0^2\tau_0^4} = -k_0''z_0$$

for some positive distance z_0, then the effective width (3.7.8) can be written as:

$$\tau_{\text{chirp}}^2 + jk_0''z = \frac{\tau_0^2}{1+\dot{\omega}_0^2\tau_0^4} + jk_0''(z - z_0) \tag{3.7.9}$$

and as z increases over the interval $0 \le z \le z_0$, the pulse width will be getting narrower, becoming the narrowest at $z = z_0$. Beyond, $z > z_0$, the pulse width will start increasing again. Thus, the initial chirping and the chirping due to propagation cancel each other at $z = z_0$. Some dispersion compensation methods are based on this effect.

3.8 Dispersion Compensation

The filtering effect of the propagation medium is represented in the frequency domain by $\hat{F}(z,\omega) = G(z,\omega)\hat{F}(0,\omega)$, where the transfer function $G(z,\omega)$ is given by Eq. (3.5.20).

To counteract the effect of spreading, a compensation filter $H_{\text{comp}}(\omega)$ may be inserted at the end of the propagation medium as shown in Fig. 3.8.1 that effectively equalizes the propagation response, up to a prescribed delay t_d, that is,

$$G(z,\omega)H_{\text{comp}}(\omega) = e^{-j\omega t_d} \quad\Rightarrow\quad \boxed{H_{\text{comp}}(\omega) = \frac{e^{-j\omega t_d}}{G(z,\omega)}} \tag{3.8.1}$$

The overall compensated output will be the input delayed by t_d, that is, $F_{\text{comp}}(z,t) = F(0,t-t_d)$. For example, if the delay is chosen to be the group delay $t_d = t_g = k_0'z$, then, in the quadratic approximation for $G(z,\omega)$, condition (3.8.1) reads:

$$G(z,\omega)H_{\text{comp}}(\omega) = e^{-jk_0'z\omega}e^{-jk_0''z\omega^2/2}H_{\text{comp}}(\omega) = e^{-jk_0'z\omega}$$

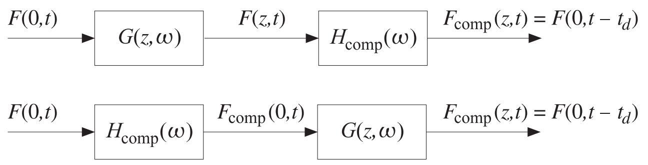

$F(0,t)$
$G(z,\omega)$
$F(z,t)$
$H_{\text{comp}}(\omega)$
$F_{\text{comp}}(z,t) = F(0,t-t_d)$
$F(0,t)$
$H_{\text{comp}}(\omega)$
$F_{\text{comp}}(0,t)$
$G(z,\omega)$
$F_{\text{comp}}(z,t) = F(0,t-t_d)$

Fig. 3.8.1 Dispersion compensation filters.

which gives for the compensation filter:

$$\boxed{H_{\text{comp}}(\omega) = e^{jk_0''z\omega^2/2}} \tag{3.8.2}$$

with impulse response:

$$h_{\text{comp}}(t) = \frac{1}{\sqrt{-2\pi jk_0''z}}\,\exp\left[\frac{t^2}{2jk_0''z}\right] \tag{3.8.3}$$

The output of the compensation filter will then agree with that of the linear approximation case, that is, it will be the input delayed as a whole by the group delay:

$$F_{\text{comp}}(z,\omega) = H_{\text{comp}}(\omega)\hat{F}(z,\omega) = H_{\text{comp}}(\omega)G(z,\omega)\hat{F}(0,\omega) = e^{-jk_0'z\omega}\hat{F}(0,\omega)$$

or, in the time domain, $F_{\text{comp}}(z,t) = F(0,t-k_0'z)$.

As shown in Fig. 3.8.1, it is possible [198] to insert the compensation filter at the input end. The pre-compensated input then suffers an equal and opposite dispersion as it propagates by a distance z, resulting in the same compensated output. As an example, an input gaussian and its pre-compensated version will be:

$$\hat{F}(0,\omega) = \sqrt{2\pi\tau_0^2}\,e^{-\tau_0^2\omega^2/2}, \quad \hat{F}_{\text{comp}}(0,\omega) = H_{\text{comp}}(\omega)\hat{F}(0,\omega) = \sqrt{2\pi\tau_0^2}\,e^{-(\tau_0^2 - jk_0''z)\omega^2/2}$$

and in the time domain:

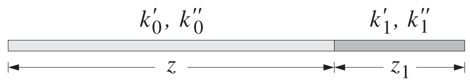
$$F(0,t) = \exp\left[-\frac{t^2}{2\tau_0^2}\right], \quad F_{\text{comp}}(0,t) = \sqrt{\frac{\tau_0^2}{\tau_0^2 - jk_0''z}}\,\exp\left[-\frac{t^2}{2(\tau_0^2 - jk_0''z)}\right]$$

This corresponds to a chirped gaussian input with a chirping parameter opposite that of Eq. (3.7.6). If the pre-compensated signal is propagated by a distance z, then its new complex-width will be, $(\tau_0^2 - jk_0''z) + jk_0''z = \tau_0^2$, and its new amplitude:

$$\sqrt{\frac{\tau_0^2}{\tau_0^2 - jk_0''z}}\,\sqrt{\frac{\tau_0^2 - jk_0''z}{(\tau_0^2 - jk_0''z) + jk_0''z}} = 1$$

thus, including the group delay, the propagated signal will be $F_{\text{comp}}(z,t) = F(0,t-k_0'z)$.

There are many ways of implementing dispersion compensation filters in optical fiber applications, such as using appropriately chirped inputs, or using fiber delay-line filters at either end, or appending a length of fiber that has equal end opposite dispersion. The latter method is one of the most widely used and is depicted below:



To see how it works, let the appended fiber have length z_1 and group delay and dispersion parameters k_1', k_1''. Then, its transfer function will be:

$$G_1(z_1,\omega) = e^{-jk_1'z_1\omega}e^{-jk_1''z_1\omega^2/2}$$

The combined transfer function of propagating through the main fiber of length z followed by z_1 will be:

$$\begin{aligned} G(z,\omega)G_1(z_1,\omega) &= e^{-jk_0'z\omega}e^{-jk_0''z\omega^2/2}e^{-jk_1'z_1\omega}e^{-jk_1''z_1\omega^2/2} \\ &= e^{-j(k_0'z+k_1'z_1)\omega}e^{-j(k_0''z+k_1''z_1)\omega^2/2} \end{aligned} \tag{3.8.4}$$

If k_1'' has the opposite sign from k_0'' and z_1 is chosen such that $k_0''z + k_1''z_1 = 0$, or, $k_1''z_1 = -k_0''z$, then the dispersion will be canceled. Thus, up to a delay, $G_1(z_1,\omega)$ acts just like the required compensation filter $H_{\text{comp}}(\omega)$. In practice, the appended fiber is manufactured to have $|k_1''| \gg |k_0''|$, so that its length will be short, $z_1 = -k_0''z/k_1'' \ll z$.

3.9 Slow, Fast, and Negative Group Velocities

The group velocity approximations of Sec. 3.5 are valid when the signal band is narrowly centered about a carrier frequency ω_0 around which the wavenumber $k(\omega)$ is a slowly-varying function of frequency to justify the Taylor series expansion (3.5.9).

The approximations are of questionable validity in spectral regions where the wavenumber, or equivalently, the refractive index $n(\omega)$, are varying rapidly with frequency, such as in the immediate vicinity of absorption or gain resonances in the propagation medium. However, even in such cases, the basic group velocity approximation, $F(z,t)= F(0,t-z/v_g)$, can be justified provided the signal bandwidth $\Delta\omega$ is sufficiently narrow and the propagation distance z is sufficiently short to minimize spreading and chirping; for example, in the gaussian case, this would require the condition $|k_0''z| \ll \tau_0^2$, or, $|k_0''z(\Delta\omega)^2| \ll 1$, as well as the condition $|\,\text{Im}(k_0)z| \ll 1$ to minimize amplitude distortions due to absorption or gain.

Because near resonances the group velocity v_g can be subluminal, superluminal, or negative, this raises the issue of how to interpret the result $F(z,t)= F(0,t-z/v_g)$. For example, if v_g is negative within a medium of thickness z, then the group delay $t_g = z/v_g$ will be negative, corresponding to a *time advance*, and the envelope's peak will appear to exit the medium before it even enters it. Indeed, experiments have demonstrated such apparently bizarre behavior [228,229,247]. As we mentioned in Sec. 3.2, this is not at odds with relativistic causality because the peaks are not necessarily causally related—only sharp signal fronts may not travel faster than c.

The gaussian pulses used in the above experiments do not have a sharp front. Their (infinitely long) forward tail can enter and exit the medium well before the peak does. Because of the spectral reshaping taking place due to the propagation medium's response $e^{-jk(\omega)z}$, the forward portion of the pulse that is already within the propagation medium, and the portion that has already exited, can get reshaped into a peak that appears to have exited before the peak of the input has entered. In fact, before the incident peak enters the medium, two additional peaks develop caused by the forward tail of the input: the one that has already exited the medium, and another one within the medium traveling *backwards* with the negative group velocity v_g. Such backward-moving peaks have been observed experimentally [275]. We clarify these remarks later on by means of the numerical example shown in Fig. 3.9.4 and elaborated further in Problem 3.10.

Next, we look at some examples that are good candidates for demonstrating the above ideas. We recall from Sec. 1.11 the following relationships between wavenumber $k = \beta - j\alpha$, refractive index $n = n_r - jn_i$, group index n_g, and dispersion coefficient k'', where all the quantities are functions of the frequency ω:

$$\begin{aligned} k &= \beta - j\alpha = \frac{\omega n}{c} = \frac{\omega(n_r - jn_i)}{c} \\ k' &= \frac{dk}{d\omega} = \frac{1}{c}\frac{d(\omega n)}{d\omega} = \frac{n_g}{c} \quad\Rightarrow\quad v_g = \frac{1}{\text{Re}(k')} = \frac{c}{\text{Re}(n_g)} \\ k'' &= \frac{d^2k}{d\omega^2} = \frac{1}{c}\frac{dn_g}{d\omega} = \frac{n_g'}{c} \end{aligned} \tag{3.9.1}$$

We consider first a single-resonance absorption or gain Lorentz medium with permittivity given by Eq. (1.9.15), that is, having susceptibility χ and refractive index n:

$$\chi = \frac{f\omega_p^2}{\omega_r^2 - \omega^2 + j\omega\gamma} \quad\Rightarrow\quad n = \sqrt{1+\chi} = \sqrt{1 + \frac{f\omega_p^2}{\omega_r^2 - \omega^2 + j\omega\gamma}} \tag{3.9.2}$$

where ω_r, γ are the resonance frequency and linewidth, and ω_p, f are the plasma frequency and oscillator strength. For an absorption medium, we will set $f = 1$, for a gain medium, $f = -1$, and for vacuum, $f = 0$. To simplify the algebra, we may use the approximation (1.11.3), that is,

$$n = \sqrt{1+\chi} \simeq 1 + \frac{1}{2}\chi = 1 + \frac{f\omega_p^2/2}{\omega_r^2 - \omega^2 + j\omega\gamma} \tag{3.9.3}$$

This approximation is fairly accurate in the numerical examples that we consider. The corresponding complex-valued group index follows from (3.9.3):

$$n_g = \frac{d(\omega n)}{d\omega} = 1 + \frac{f\omega_p^2(\omega^2 + \omega_r^2)/2}{(\omega_r^2 - \omega^2 + j\omega\gamma)^2} \tag{3.9.4}$$

with real and imaginary parts:

$$\begin{aligned} \text{Re}(n_g) &= 1 + \frac{f\omega_p^2(\omega^2 + \omega_r^2)\big[(\omega^2 - \omega_r^2)^2 - \omega^2\gamma^2\big]}{\big[(\omega^2 - \omega_r^2)^2 + \omega^2\gamma^2\big]^2} \\ \text{Im}(n_g) &= \frac{f\omega_p^2\gamma\omega(\omega^4 - \omega_r^4)}{\big[(\omega^2 - \omega_r^2)^2 + \omega^2\gamma^2\big]^2} \end{aligned} \tag{3.9.5}$$

Similarly, the dispersion coefficient $dn_g/d\omega$ is given by:

$$n_g' = \frac{dn_g}{d\omega} = \frac{f\omega_p^2(\omega^3 + 3\omega_r^2\omega - j\gamma\omega_r^2)}{(\omega_r^2 - \omega^2 + j\omega\gamma)^3} \tag{3.9.6}$$

At resonance, $\omega = \omega_r$, we find the values:

$$n = 1 - j\frac{f\omega_p^2}{2\gamma\omega_r}\,, \quad n_g = 1 - \frac{f\omega_p^2}{\gamma^2} \tag{3.9.7}$$

For an absorption medium ($f = 1$), if $\omega_p < \gamma$, the group index will be $0 < n_g < 1$, resulting into a superluminal group velocity $v_g = c/n_g > c$, but if $\gamma < \omega_p$, which is the more typical case, then the group index will become negative, resulting into a negative $v_g = c/n_g < 0$. This is illustrated in the top row of graphs of Fig. 3.9.1. On the other hand, for a gain medium ($f = -1$), the group index is always $n_g > 1$ at resonance, resulting into a subluminal group velocity $v_g = c/n_g < c$. This is illustrated in the middle and bottom rows of graphs of Fig. 3.9.1.

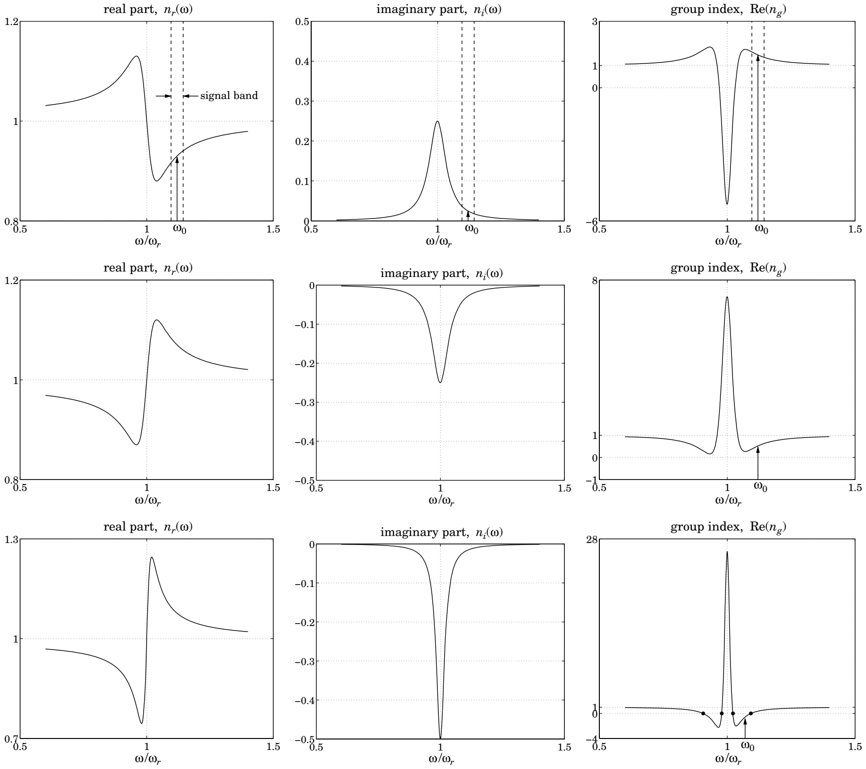


Fig. 3.9.1 Slow, fast, and negative group velocities (at off resonance).

Fig. 3.9.1 plots $n(\omega) = n_r(\omega) - jn_i(\omega)$ and $\text{Re}[n_g(\omega)]$ versus ω, evaluated using Eqs. (3.9.3) and (3.9.4), with the frequency axis normalized in units of ω/ω_r. The following values of the parameters were used (with arbitrary frequency units):

$$\begin{array}{lllll}
\text{(top row)} & f = +1\,, & \omega_p = 1\,, & \omega_r = 5\,, & \gamma = 0.4 \\
\text{(middle row)} & f = -1\,, & \omega_p = 1\,, & \omega_r = 5\,, & \gamma = 0.4 \\
\text{(bottom row)} & f = -1\,, & \omega_p = 1\,, & \omega_r = 5\,, & \gamma = 0.2
\end{array}$$

The calculated values of n, n_g at resonance were:

$$\begin{array}{llll}
\text{(top)} & \omega = \omega_r\,, & n = 1 - 0.25j\,, & n_g = -5.25 \\
\text{(middle)} & \omega = \omega_r\,, & n = 1 + 0.25j\,, & n_g = 7.25 \\
\text{(bottom)} & \omega = \omega_r\,, & n = 1 + 0.5j\,, & n_g = 26
\end{array}$$

Operating at resonance is not a good idea because of the fairly substantial amounts of attenuation or gain arising from the imaginary part n_i of the refractive index, which would cause amplitude distortions in the signal as it propagates.

A better operating frequency band is at off resonance where the attenuation or gain are lower [234]. The top row of Fig. 3.9.1 shows such a band centered at a frequency ω_0 on the right wing of the resonance, with a narrow enough bandwidth to justify the Taylor series expansion (3.5.9). The group velocity behavior is essentially the reverse of that at resonance, that is, v_g becomes subluminal for the absorption medium, and superluminal or negative for the gain medium. The carrier frequency ω_0 and the calculated values of n, n_g at $\omega = \omega_0$ were as follows:

$$\begin{array}{llll}
\text{(top, slow)} & \omega_0/\omega_r = 1.12\,, & n = 0.93 - 0.02j\,, & n_g = 1.48 + 0.39j \\
\text{(middle, fast)} & \omega_0/\omega_r = 1.12\,, & n = 1.07 + 0.02j\,, & n_g = 0.52 - 0.39j \\
\text{(bottom, negative)} & \omega_0/\omega_r = 1.07\,, & n = 1.13 + 0.04j\,, & n_g = -0.58 - 1.02j
\end{array}$$

We note the sign and magnitude of $\text{Re}(n_g)$ and the substantially smaller values of the imaginary part n_i. For the middle graph, the group index remains in the interval $0 < \text{Re}(n_g) < 1$, and hence $v_g > c$, for all values of the frequency in the right wing of the resonance.

In order to get negative values for $\text{Re}(n_g)$ and for v_g, the linewidth γ must be reduced. As can be seen in the bottom row of graphs, $\text{Re}(n_g)$ becomes negative over a small range of frequencies to the right and left of the resonance. The edge frequencies can be calculated from the zero-crossings of $\text{Re}(n_g)$ and are shown on the graph. For the given parameter values, they were found to be (in units of ω/ω_r):

$$[0.9058,\ 0.9784]\,, \quad [1.0221,\ 1.0928]$$

The chosen value of $\omega_0/\omega_r = 1.07$ falls inside the right interval.

Another way of demonstrating slow, fast, or negative group velocities with low attenuation or gain, which has been used in practice, is to operate at a frequency band that lies between two nearby absorption or gain lines.

Some examples are shown in Fig. 3.9.2. The top row of graphs depicts the case of two nearby absorption lines. In the band between the lines, the refractive index exhibits normal dispersion. Exactly at midpoint, the attenuation is minimal and the real part n_r has a steep slope that causes a large group index, $\text{Re}(n_g) \gg 1$, and hence a small positive group velocity $0 < v_g \ll c$. In experiments, very sharp slopes have been

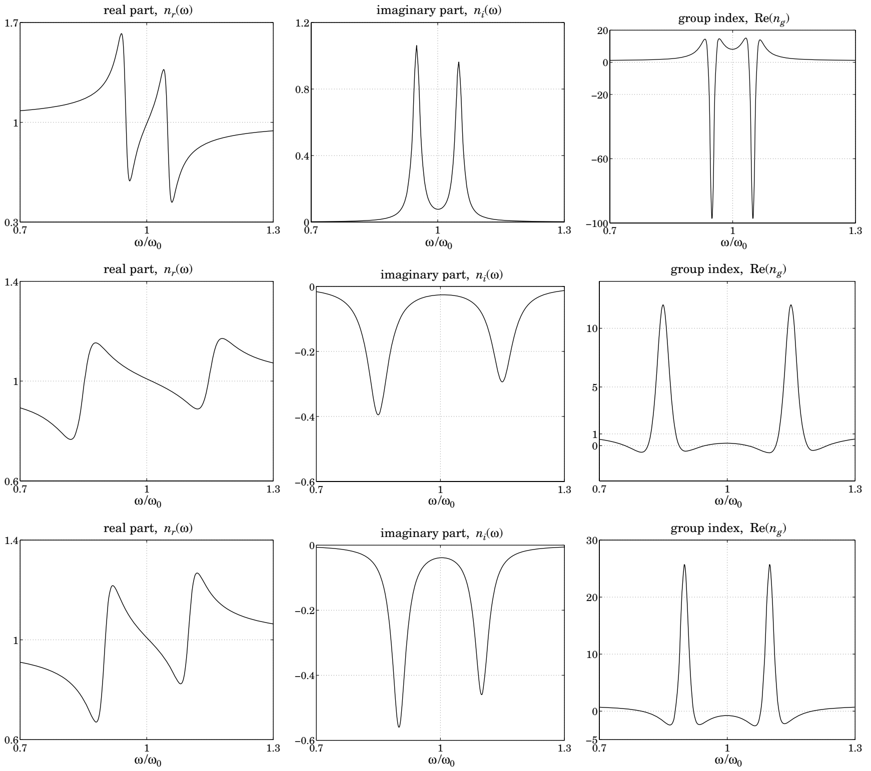


Fig. 3.9.2 Slow, fast, and negative group velocities (halfway between resonances).

achieved through the use of the so-called “electromagnetically induced transparency,” resulting into extremely slow group velocities of the order of tens of m/sec [289].

The middle row of graphs depicts two nearby gain lines [235] with a small gain at midpoint and a real part n_r that has a negative slope resulting into a group index $0 < \mathrm{Re}(n_g) < 1$, and a superluminal group velocity $v_g > c$.

Choosing more closely separated peaks in the third row of graphs, has the effect of increasing the negative slope of n_r, thus causing the group index to become negative at midpoint, $\mathrm{Re}(n_g) < 0$, resulting in negative group velocity, $v_g < 0$. Experiments demonstrating this behavior have received a lot of attention [247].

The following expressions were used in Fig. 3.9.2 for the refractive and group indices,

with $f = 1$ for the absorption case, and $f = -1$ for the gain case:

$$\begin{aligned} n &= 1 + \frac{f\omega_p^2/2}{\omega_1^2 - \omega^2 + j\omega\gamma} + \frac{f\omega_p^2/2}{\omega_2^2 - \omega^2 + j\omega\gamma} \\ n_g &= 1 + \frac{f\omega_p^2(\omega^2+\omega_1^2)/2}{(\omega_1^2 - \omega^2 + j\omega\gamma)^2} + \frac{f\omega_p^2(\omega^2+\omega_2^2)/2}{(\omega_2^2 - \omega^2 + j\omega\gamma)^2} \end{aligned} \tag{3.9.8}$$

The two peaks were symmetrically placed about the midpoint frequency ω_0, that is, at $\omega_1 = \omega_0 - \Delta$ and $\omega_2 = \omega_0 + \Delta$, and a common linewidth γ was chosen. The particular numerical values used in this graph were:

$$\begin{array}{llllll} \text{(top, slow)} & f = +1\,, & \omega_p = 1\,, & \omega_0 = 5\,, & \Delta = 0.25\,, & \gamma = 0.1 \\ \text{(middle, fast)} & f = -1\,, & \omega_p = 1\,, & \omega_0 = 5\,, & \Delta = 0.75\,, & \gamma = 0.3 \\ \text{(bottom, negative)} & f = -1\,, & \omega_p = 1\,, & \omega_0 = 5\,, & \Delta = 0.50\,, & \gamma = 0.2 \end{array}$$

resulting in the following values for n and n_g:

$$\begin{array}{lll} \text{(top, slow)} & n = 0.991 - 0.077j\,, & n_g = 8.104 + 0.063j \\ \text{(middle, fast)} & n = 1.009 + 0.026j\,, & n_g = 0.208 - 0.021j \\ \text{(bottom, negative)} & n = 1.009 + 0.039j\,, & n_g = -0.778 - 0.032j \end{array}$$

Next, we look at an example of a gaussian pulse propagating through a medium with negative group velocity. We consider a single-resonance gain medium and operating frequency band similar to that shown in the bottom row of graphs in Fig. 3.9.1. This example is variation of that discussed in [234]. Fig. 3.9.3 shows the geometry.

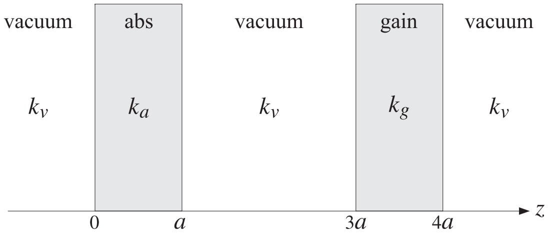


Fig. 3.9.3 Absorption and gain media separated by vacuum.

The gaussian pulse begins in vacuum on the left, then it enters an absorbing medium of thickness a in which it propagates with a slow group velocity suffering a modest amount of attenuation. It then enters a vacuum region of width $2a$, followed by a gain medium of thickness a in which it propagates with negative group velocity suffering a moderate amount gain, and finally it exits into vacuum.

The attenuation and gain are adjusted to compensate each other, so that the final output vacuum pulse is identical to the input.

The wavenumbers k_v, k_a, k_g, in vacuum, the absorption and gain media are calculated from Eqs. (3.9.1)–(3.9.6) with $f = 0, +1, -1$, respectively. The analytical and

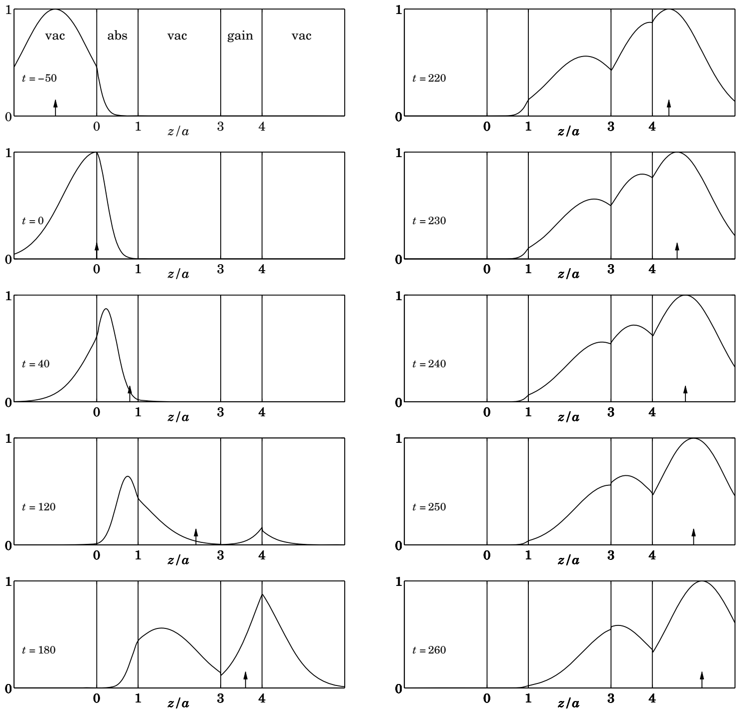


Fig. 3.9.4 Snapshots of pulse propagating through regions of different group velocities.

numerical details for this example are outlined in Problem 3.10. Fig. 3.9.4 shows a series of snapshots. The short vertical arrow on the horizontal axis represents the position of the peak of an equivalent pulse propagating in vacuum.

At $t = -50$ (in units such that $c = 1$), the forward tail of the gaussian pulse has already entered the absorbing medium. Between $0 \le t \le 120$, the peak of the pulse has entered the absorbing medium and is being attenuated as it propagates while it lags behind the equivalent vacuum pulse because $v_g < c$.

At $t = 120$, while the peak is still in the absorbing medium, the forward tail has passed through the middle vacuum region and has already entered into the gain medium where it begins to get amplified. At $t = 180$, the peak has moved into the middle vacuum region, but the forward tail has been sufficiently amplified by the gain medium and is beginning to form a peak whose tail has already exited into the rightmost vacuum region.

At $t = 220$, the peak is still within the middle vacuum region, but the output peak has already exited into the right, while another peak has formed at the right side of the gain medium and begins to move backwards with the negative group velocity, $v_g < 0$. Meanwhile, the output peak has caught up with the equivalent vacuum peak.

Between $230 \le t \le 260$, the peak within the gain medium continues to move backwards while the output vacuum peak moves to the right. As we mentioned earlier, such output peaks that have exited before the input peaks have entered the gain medium, including the backward moving peaks, have been observed experimentally [275].

A MATLAB movie of this example may be seen by running the file `grvmovie1.m` in the `movies` subdirectory of the ewa toolbox. See also the movie `grvmovie2.m` in which the carrier frequency has been increased and corresponds to a superluminal group velocity ($v_g > c$) for the gain medium. In this case, which is also described in Problem 3.10, all the peaks are moving forward.

3.10 Chirp Radar and Pulse Compression

Pulse Radar Requirements

The chirping and dispersion compensation concepts discussed in the previous sections are applicable also to chirp radar systems. Here, we give a brief introduction to the main ideas [320] and the need for pulse compression.

In radar, the propagation medium is assumed to be non-dispersive (e.g., air), hence, it introduces only a propagation delay. Chirping is used to increase the bandwidth of the transmitted radar pulses, while keeping their time-duration long. The received pulses are processed by a dispersion compensation filter that cancels the frequency dispersion introduced by chirping and results in a time-compressed pulse. The basic system is shown in Fig. 3.10.1. The technique effectively combines the benefits of a long-duration pulse (improved detectability and Doppler resolution) with those of a broadband pulse (improved range resolution.)

A typical pulsed radar sends out sinusoidal pulses of some finite duration of, say, T seconds. A pulse reflected from a stationary target at a distance R returns back at the radar attenuated and with an overall round-trip delay of $t_d = 2R/c$ seconds. The range R is determined from the delay t_d. An uncertainty in measuring t_d from two nearby targets translates into an uncertainty in the range, $\Delta R = c(\Delta t_d)/2$. Because the pulse has duration T, the uncertainty in t_d will be $\Delta t_d = T$, and the uncertainty in the range, $\Delta R = cT/2$. Thus, to improve the range resolution, a short pulse duration T must be used.

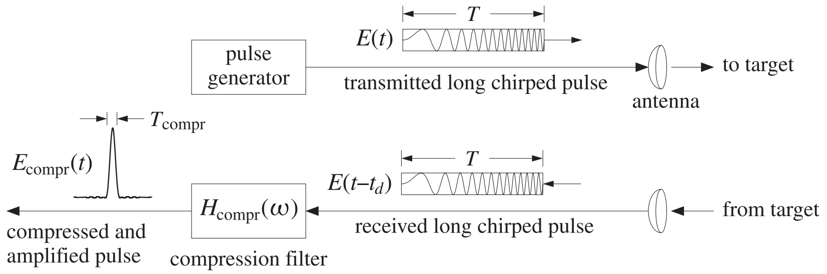


Fig. 3.10.1 Chirp radar system.

On the other hand, the detectability of the received pulse requires a certain minimum value of the signal-to-noise ratio (SNR), which in turn, requires a large value of T. The SNR at the receiver is given by

$$\text{SNR} = \frac{\mathcal{E}_{\text{rec}}}{N_0} = \frac{P_{\text{rec}}T}{N_0}$$

where P_{rec} and $\mathcal{E}_{\text{rec}} = P_{\text{rec}}T$ denote the power and energy of the received pulse, and N_0 is the noise power spectral density given in terms of the effective noise temperature T_e of the receiver by $N_0 = kT_e$ (as discussed in greater detail in Sec. 15.7). It follows from the radar equation (15.11.4) of Sec. 15.11, that the received power P_{rec} is proportional to the transmitter power P_{tr} and inversely proportional to the fourth power of the distance R. Thus, to keep the SNR at detectable levels for large distances, a large transmitter power and corresponding pulse energy $\mathcal{E}_{\text{tr}} = P_{\text{tr}}T$ must be used. This can be achieved by increasing T, while keeping P_{tr} at manageable levels.

The Doppler velocity resolution, similarly, improves with increasing T. The Doppler frequency shift for a target moving at a radial velocity v is $f_d = 2f_0 v/c$, where f_0 is the carrier frequency. We will see below that the uncertainty in f_d is given roughly by $\Delta f_d = 1/T$. Thus, the uncertainty in speed will be $\Delta v = c(\Delta f_d)/2f_0 = c/(2f_0 T)$.

The simultaneous conflicting requirements of a short duration T to improve the resolution in range, and a large duration T to improve the detectability of distant targets and Doppler resolution, can be realized by sending out a pulse that has both a long duration T and a very large bandwidth of, say, B Hertz, such that $BT \gg 1$. Upon reception, the received pulse can be compressed with the help of a compression filter to the much shorter duration of $T_{\text{compr}} = 1/B$ seconds, which satisfies $T_{\text{compr}} = 1/B \ll T$. The improvement in range resolution will be then $\Delta R = cT_{\text{compr}}/2 = c/2B$.

In summary, the following formulas capture the tradeoffs among the three requirements of detectability, range resolution, and Doppler resolution:

$$\text{SNR} = \frac{\mathcal{E}_{\text{rec}}}{N_0} = \frac{P_{\text{rec}}T}{N_0}, \quad \Delta R = \frac{c}{2B}, \quad \Delta v = \frac{c}{2f_0 T} \tag{3.10.1}$$

For example, to achieve a 30-meter range resolution and a 50 m/s (180 km/hr) velocity resolution at a 3-GHz carrier frequency, would require $B = 5$ MHz and $T = 1$ msec, resulting in the large time-bandwidth product of $BT = 5000$.

Such large time-bandwidth products cannot be achieved with plain sinusoidal pulses. For example, an ordinary, unchirped, sinusoidal rectangular pulse of duration of T seconds has an effective bandwidth of $B = 1/T$ Hertz, and hence, $BT = 1$. This follows from the Fourier transform pair:

$$E(t) = \text{rect}\Big(\frac{t}{T}\Big)e^{j\omega_0 t} \quad \Leftrightarrow \quad \hat{E}(\omega) = T\,\frac{\sin\big((\omega-\omega_0)T/2\big)}{(\omega-\omega_0)T/2} \tag{3.10.2}$$

where $\text{rect}(x)$ is the rectangular pulse defined with the help of the unit step $u(x)$:

$$\text{rect}(x) = u(x+0.5) - u(x-0.5) = \begin{cases} 1, & \text{if } |x| < 0.5 \\ 0, & \text{if } |x| > 0.5 \end{cases}$$

It follows from (3.10.2) that the 3-dB width of the spectrum is $\Delta\omega = 0.886(2\pi)/T$, or in Hz, $\Delta f = 0.886/T$, and similarly, the quantity $\Delta f = 1/T$ represents the 4-dB width. Thus, the effective bandwidth of the rectangular pulse is $1/T$.

Linear FM Signals

It is possible, nevertheless, to have a waveform whose envelope has an arbitrary duration T while its spectrum has an arbitrary width B, at least in an approximate sense. The key idea in accomplishing this is to have the *instantaneous frequency* of the signal vary—during the duration T of the envelope—over a set of values that *span* the desired bandwidth B. Such time variation of the instantaneous frequency translates in the frequency domain to a spectrum of effective width B.

The simplest realization of this idea is through *linear FM*, or chirping, that corresponds to a linearly varying instantaneous frequency. More complicated schemes exist that use nonlinear time variations, or, using phase-coding in which the instantaneous phase of the signal changes by specified amounts during the duration T in such a way as to broaden the spectrum. A chirped pulse is given by:

$$\boxed{E(t) = F(t)\,e^{j\omega_0 t + j\dot{\omega}_0 t^2/2}} \tag{3.10.3}$$

where $F(t)$ is an arbitrary envelope with an effective duration T, defined for example over the time interval $-T/2 \le t \le T/2$. The envelope $F(t)$ can be specified either in the time domain or in the frequency domain by means of its spectrum $\hat{F}(\omega)$:

$$\hat{F}(\omega) = \int_{-\infty}^{\infty} F(t)\,e^{-j\omega t}\,dt \quad \Leftrightarrow \quad F(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{F}(\omega)\,e^{j\omega t}\,d\omega \tag{3.10.4}$$

Typically, $F(t)$ is real-valued and therefore, the instantaneous frequency of (3.10.3) is $\omega(t) = \dot{\theta}(t) = \omega_0 + \dot{\omega}_0 t$. During the time interval $-T/2 \le t \le T/2$, it varies over the band $\omega_0 - \dot{\omega}_0 T/2 \le \omega(t) \le \omega_0 + \dot{\omega}_0 T/2$, (we are assuming here that $\dot{\omega}_0 > 0$.) Hence, it has an effective total bandwidth:

$$\Omega = \dot{\omega}_0 T, \quad \text{or, in units of Hz}, \quad B = \frac{\Omega}{2\pi} = \frac{\dot{\omega}_0 T}{2\pi} \tag{3.10.5}$$

Thus, given T and B, the chirping parameter can be chosen to be $\dot{\omega}_0 = 2\pi B/T$. We will look at some examples of $F(t)$ shortly and confirm that the spectrum of the chirped signal $E(t)$ is effectively confined in the band $|f - f_0| \le B/2$. But first, we determine the compression filter.

Pulse Compression Filter

The received signal reflected from a target is an attenuated and delayed copy of the transmitted signal $E(t)$, that is,

$$E_{\text{rec}}(t) = aE(t - t_d) = aF(t - t_d)\,e^{j\omega_0(t-t_d) + j\dot{\omega}_0(t-t_d)^2/2} \tag{3.10.6}$$

where a is an attenuation factor determined from the radar equation to be the ratio of the received to the transmitted powers: $a^2 = P_{\text{rec}}/P_{\text{tr}}$.

If the target is moving with a radial velocity v towards the radar, there will be a Doppler shift by $\omega_d = 2v\omega_0/c$. Although this shift affects all the frequency components, that is, $\omega \to \omega + \omega_d$, it is common to make the so-called narrowband approximation in which only the carrier frequency is shifted $\omega_0 \to \omega_0 + \omega_d$. This is justified for radar signals because, even though the bandwidth Ω is wide, it is still only a small fraction (typically one percent) of the carrier frequency, that is, $\Omega \ll \omega_0$. Thus, the received signal from a moving target is taken to be:

$$E_{\text{rec}}(t)=aE(t-t_d)\,e^{j\omega_d(t-t_d)} = aF(t-t_d)\,e^{j(\omega_0+\omega_d)(t-t_d)+j\dot{\omega}_0(t-t_d)^2/2} \tag{3.10.7}$$

To simplify the notation, we will ignore the attenuation factor and the delay, which can be restored at will later, and take the received signal to be:

$$\boxed{E_{\text{rec}}(t)=E(t)\,e^{j\omega_d t} = F(t)\,e^{j(\omega_0+\omega_d)t+j\dot{\omega}_0 t^2/2}} \tag{3.10.8}$$

This signal is then processed by a pulse compression filter that will compress the waveform to a shorter duration. To determine the specifications of the compression filter, we consider the unrealizable case of a signal that has infinite duration and infinite bandwidth defined by $F(t)=1$, for $-\infty<t<\infty$. For now, we will ignore the Doppler shift so that $E_{\text{rec}}(t)=E(t)$. Using Eq. (3.5.18), the chirped signal and its spectrum are:

$$E(t)=e^{j\omega_0 t+j\dot{\omega}_0 t^2/2} \quad\Leftrightarrow\quad \hat{E}(\omega)=\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,e^{-j(\omega-\omega_0)^2/2\dot{\omega}_0} \tag{3.10.9}$$

Clearly, the magnitude spectrum is constant and has infinite extent spanning the entire frequency axis. The compression filter must equalize the quadratic phase spectrum of the signal, that is, it must have the opposite phase:

$$\boxed{H_{\text{compr}}(\omega)=e^{j(\omega-\omega_0)^2/2\dot{\omega}_0}} \quad \text{(pulse compression filter)} \tag{3.10.10}$$

The corresponding impulse response is the inverse Fourier transform of Eq. (3.10.10):

$$\boxed{h_{\text{compr}}(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t-j\dot{\omega}_0 t^2/2}} \quad \text{(pulse compression filter)} \tag{3.10.11}$$

The resulting output spectrum for the input (3.10.9) will be:

$$\hat{E}_{\text{compr}}(\omega)=H_{\text{compr}}(\omega)\hat{E}(\omega)=\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,e^{-j(\omega-\omega_0)^2/2\dot{\omega}_0}\cdot e^{j(\omega-\omega_0)^2/2\dot{\omega}_0}=\sqrt{\frac{2\pi j}{\dot{\omega}_0}}$$

that is, a constant for all ω. Hence, the input signal gets compressed into a Dirac delta:

$$E_{\text{compr}}(t)=\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,\delta(t) \tag{3.10.12}$$

When the envelope $F(t)$ is a finite-duration signal, the resulting spectrum of the chirped signal $E(t)$ still retains the essential quadratic phase of Eq. (3.10.9), and therefore, the compression filter will still be given by Eq. (3.10.10) for all choices of $F(t)$. Using the stationary-phase approximation, Problem 3.17 shows that the quadratic phase is a general property. The group delay of this filter is given by Eq. (3.2.1):

$$t_g=-\frac{d}{d\omega}\left[\frac{(\omega-\omega_0)^2}{2\dot{\omega}_0}\right]=-\frac{\omega-\omega_0}{\dot{\omega}_0}=-\frac{2\pi(f-f_0)}{2\pi B/T}=-T\,\frac{f-f_0}{B}$$

As the frequency $(f-f_0)$ increases from $-B/2$ to $B/2$, the group delay *decreases* from $T/2$ to $-T/2$, that is, the lower frequency components, which occur earlier in the chirped pulse, suffer a longer delay through the filter. Similarly, the high frequency components, which occur later in the pulse, suffer a shorter delay, the overall effect being the time compression of the pulse.

It is useful to demodulate the sinusoidal carrier $e^{j\omega_0 t}$ and write $h_{\text{compr}}(t)=e^{j\omega_0 t}g(t)$ and $H_{\text{compr}}(\omega)=G(\omega-\omega_0)$, where the demodulated "baseband" filter, which is known as a *quadrature-phase filter*, is defined by:

$$\boxed{g(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{-j\dot{\omega}_0 t^2/2}\,,\quad G(\omega)=e^{j\omega^2/2\dot{\omega}_0}} \quad \text{(quadratic phase filter)} \tag{3.10.13}$$

For an arbitrary envelope $F(t)$, one can derive the following fundamental result that relates the *output* of the compression filter (3.10.11) to the *Fourier transform*, $\hat{F}(\omega)$, of the envelope, when the input is $E(t)=F(t)\,e^{j\omega_0 t+j\dot{\omega}_0 t^2/2}$:

$$\boxed{E_{\text{compr}}(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t-j\dot{\omega}_0 t^2/2}\,\hat{F}(-\dot{\omega}_0 t)} \tag{3.10.14}$$

This result belongs to a family of so-called "chirp transforms" or "Fresnel transforms" that find application in optics, the diffraction effects of lenses [1160], and in other areas of signal processing, such as for example, the "chirp z-transform" [47]. To show Eq. (3.10.14), we use the convolutional definition for the filter output:

$$\begin{aligned}
E_{\text{compr}}(t) &= \int_{-\infty}^{\infty} h_{\text{compr}}(t-t')\,E(t')\,dt' \\
&= \sqrt{\frac{j\dot{\omega}_0}{2\pi}}\int_{-\infty}^{\infty} e^{j\omega_0(t-t')-j\dot{\omega}_0(t-t')^2/2}\,F(t')\,e^{j\omega_0 t'+j\dot{\omega}_0 t'^2/2}\,dt' \\
&= \sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t-j\dot{\omega}_0 t^2/2}\int_{-\infty}^{\infty} F(t')\,e^{j(\dot{\omega}_0 t)t'}\,dt'
\end{aligned}$$

where the last integral factor is recognized as $\hat{F}(-\dot{\omega}_0 t)$. As an example, Eq. (3.10.12) can be derived immediately by noting that $F(t)=1$ has the Fourier transform $\hat{F}(\omega)=2\pi\delta(\omega)$, and therefore, using Eq. (3.10.14), we have:

$$E_{\text{compr}}(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t-j\dot{\omega}_0 t^2/2}\,2\pi\delta(-\dot{\omega}_0 t)=\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,\delta(t)$$

where we used the property $\delta(-\dot{\omega}_0 t)=\delta(\dot{\omega}_0 t)=\delta(t)/\dot{\omega}_0$ and set $t=0$ in the exponentials.

The property (3.10.14) is shown pictorially in Fig. 3.10.2. This arrangement can also be thought of as a real-time spectrum analyzer of the input envelope $F(t)$.

In order to remove the chirping factor $e^{-j\dot{\omega}_0 t^2/2}$, one can prefilter $F(t)$ with the baseband filter $G(\omega)$ and then apply the above result to its output. This leads to a modified compressed output given by:

$$\bar{E}_{\text{compr}}(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t}\,\hat{F}(-\dot{\omega}_0 t) \tag{3.10.15}$$

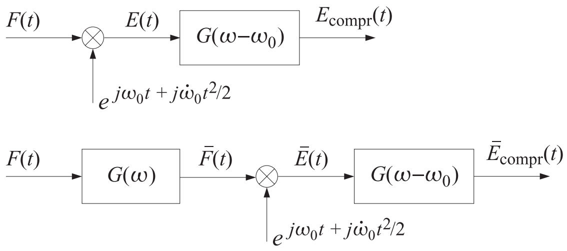


Fig. 3.10.2 Pulse compression filter.

Fig. 3.10.2 also depicts this property. To show it, we note the identity:

$$e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,\hat F(-\dot\omega_0 t) = e^{j\omega_0 t}\left[e^{-j\omega^2/2\dot\omega_0}\,\hat F(\omega)\right]_{\omega=-\dot\omega_0 t}$$

Thus, if in this expression $\hat F(\omega)$ is replaced by its prefiltered version $G(\omega)\hat F(\omega)$, then the quadratic phase factor will be canceled leaving only $\hat F(\omega)$.

For a moving target, the envelope $F(t)$ is replaced by $F(t)\,e^{j\omega_d t}$, and $\hat F(\omega)$ is replaced by $\hat F(\omega-\omega_d)$, and similarly, $\hat F(-\dot\omega_0 t)$ is replaced by $\hat F(-\dot\omega_0 t-\omega_d)$. Thus, Eq. (3.10.14) is modified as follows:

$$\boxed{E_{\rm compr}(t) = \sqrt{\frac{j\dot\omega_0}{2\pi}}\,e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,\hat F\big(-(\omega_d+\dot\omega_0 t)\big)} \tag{3.10.16}$$

Chirped Rectangular Pulse

Next, we discuss the practical case of a rectangular envelope of duration T:

$$F(t) = \mathrm{rect}\left(\frac{t}{T}\right) \quad\Rightarrow\quad E(t) = \mathrm{rect}\left(\frac{t}{T}\right)e^{j\omega_0 t + j\dot\omega_0 t^2/2} \tag{3.10.17}$$

From Eq. (3.10.2), the Fourier transform of $F(t)$ is,

$$\hat F(\omega) = T\,\frac{\sin(\omega T/2)}{\omega T/2}$$

Therefore, the output of the compression filter will be:

$$E_{\rm compr}(t) = \sqrt{\frac{j\dot\omega_0}{2\pi}}\,e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,\hat F(-\dot\omega_0 t) = \sqrt{\frac{j\dot\omega_0}{2\pi}}\,e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,T\,\frac{\sin(-\dot\omega_0 tT/2)}{-\dot\omega_0 tT/2}$$

Noting that $\dot\omega_0 T = \Omega = 2\pi B$ and that $\sqrt{j\dot\omega_0 T^2/2\pi} = \sqrt{jBT}$, we obtain:

$$\boxed{E_{\rm compr}(t) = \sqrt{jBT}\,e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,\frac{\sin(\pi Bt)}{\pi Bt}} \tag{3.10.18}$$

The sinc-function envelope $\sin(\pi Bt)/\pi Bt$ has an effective compressed width of $T_{\rm compr} = 1/B$ measured at the 4-dB level. Moreover, the height of the peak is boosted by a factor of $\sqrt{BT}$.

Fig. 3.10.3 shows a numerical example with the parameter values $T = 30$ and $B = 4$ (in arbitrary units), and $\omega_0 = 0$. The left graph plots the real part of $E(t)$ of Eq. (3.10.17). The right graph is the real part of Eq. (3.10.18), where because of the factor $\sqrt{j}$, the peak reaches the maximum value of $\sqrt{BT}/\sqrt{2}$.

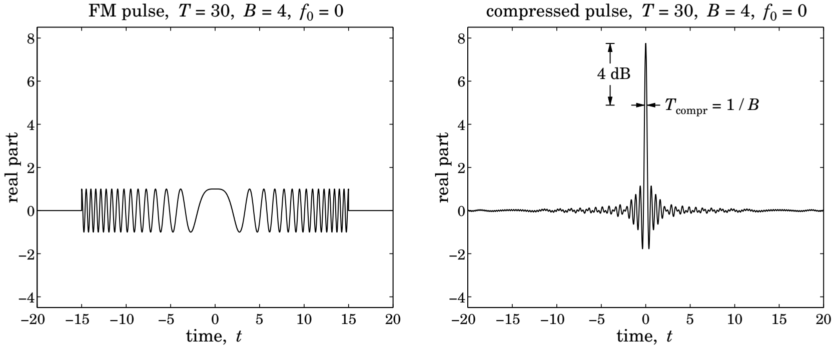


Fig. 3.10.3 FM pulse and its compressed version, with $T = 30$, $B = 4$, $f_0 = 0$.

We may also determine the Fourier transform of $E(t)$ of Eq. (3.10.17) and verify that it is primarily confined in the band $|f-f_0| \le B/2$. We have:

$$\hat E(\omega) = \int_{-\infty}^{\infty} E(t)\,e^{-j\omega t}\,dt = \int_{-T/2}^{T/2} e^{j\omega_0 t + j\dot\omega_0 t^2/2}e^{-j\omega t}\,dt$$

After changing variables from t to $u = \sqrt{\dot\omega_0/\pi}\,[t-(\omega-\omega_0)/\dot\omega_0]$, this integral can be reduced to the complex Fresnel integral $\mathcal F(x) = C(x) - jS(x) = \int_0^x e^{-j\pi u^2/2}\,du$ discussed in greater detail in Appendix F. The resulting spectrum then takes the form:

$$\hat E(\omega) = \sqrt{\frac{\pi}{\dot\omega_0}}\,e^{-j(\omega-\omega_0)^2/2\dot\omega_0}\,\big[\mathcal F(w_+) - \mathcal F(w_-)\big]^*$$

which can be written in the normalized form:

$$\hat E(\omega) = \sqrt{\frac{2\pi j}{\dot\omega_0}}\,e^{-j(\omega-\omega_0)^2/2\dot\omega_0}\,D^*(\omega)\,,\qquad D(\omega) = \frac{\mathcal F(w_+) - \mathcal F(w_-)}{1-j} \tag{3.10.19}$$

where $w_\pm$ are defined by:

$$w_\pm = \sqrt{\frac{\dot\omega_0}{\pi}}\left(\pm\frac{T}{2} - \frac{\omega-\omega_0}{\dot\omega_0}\right) = \sqrt{2BT}\left(\pm\frac{1}{2} - \frac{f-f_0}{B}\right) \tag{3.10.20}$$

Eq. (3.10.19) has the expected quadratic phase term and differs from (3.10.9) by the factor $D^*(\omega)$. This factor has a magnitude that is effectively confined within the ideal band $|f-f_0| \le B/2$ and a phase that remains almost zero within the same band, with both of these properties improving with increasing time-bandwidth product BT.[†] Thus, the choice for the compression filter that was made on the basis of the quadratic phase term is justified.

[†] The denominator $(1-j)$ in $D(\omega)$ is due to the asymptotic value of $\mathcal F(\infty) = (1-j)/2$.

Fig. 3.10.4 displays the spectrum $\hat{E}(\omega)$ for the values $T = 30$ and $B = 4$, and $\omega_0 = 0$. The left and right graphs plot the magnitude and phase of the quantity $D^*(\omega)$. For comparison, the spectrum of an ordinary, unchirped, pulse of the same duration $T = 30$, given by Eq. (3.10.2), is also shown on the magnitude graph. The Fresnel functions were evaluated with the help of the MATLAB function `fcs.m` of Appendix F. The ripples that appear in the magnitude and phase are due to the Fresnel functions.

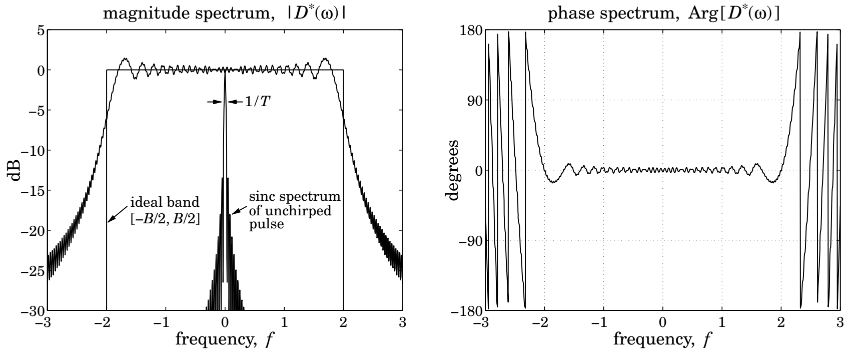


Fig. 3.10.4 Frequency spectrum of FM pulse, with $T = 30$, $B = 4$, $f_0 = 0$.

Doppler Ambiguity

For a moving target causing a Doppler shift ω_d, the output will be given by Eq. (3.10.16), which for the rectangular pulse gives:

$$E_{\text{compr}}(t)=\sqrt{\frac{j\dot\omega_0}{2\pi}}\,e^{j\omega_0 t-j\dot\omega_0 t^2/2}\,T\,\frac{\sin\big((\omega_d+\dot\omega_0 t)T/2\big)}{(\omega_d+\dot\omega_0 t)T/2}$$

Noting that $(\omega_d+\dot\omega_0 t)T = 2\pi(f_dT+Bt)$, and replacing t by $t-t_d$ to restore the actual delay of arrival of the received pulse, we obtain:

$$E_{\text{compr}}(t,f_d)=\sqrt{jBT}\,e^{j\omega_0(t-t_d)-j\dot\omega_0(t-t_d)^2/2}\,\frac{\sin\big[\pi\big(f_dT+B(t-t_d)\big)\big]}{\pi\big(f_dT+B(t-t_d)\big)} \tag{3.10.21}$$

It is seen that the peak of the pulse no longer takes place at $t = t_d$, but rather at the shifted time $f_dT+B(t-t_d)=0$, or, $t=t_d-f_dT/B$, resulting in a potential ambiguity in the range. Eq. (3.10.21) is an example of an *ambiguity function* commonly used in radar to quantify the simultaneous uncertainty in range and Doppler. Setting $t = t_d$, we find:

$$E_{\text{compr}}(t_d,f_d)=\sqrt{jBT}\,\frac{\sin(\pi f_dT)}{\pi f_dT} \tag{3.10.22}$$

which shows that the Doppler resolution is $\Delta f_d = 1/T$, as we discussed at the beginning.

Sidelobe Reduction

Although the filter output (3.10.18) is highly compressed, it has significant sidelobes that are approximately 13 dB down from the main lobe. Such sidelobes, referred to as "range sidelobes," can mask the presence of small nearby targets.

The sidelobes can be suppressed using windowing, which can be applied either in the time domain or the frequency domain. To reduce sidelobes in one domain (frequency or time), one must apply windowing to the conjugate domain (time or frequency).

Because the compressed output envelope is the Fourier transform $\hat{F}(\omega)$ evaluated at $\omega = -\dot\omega_0 t$, the sidelobes can be suppressed by applying a time window $w(t)$ of length T to the envelope, that is, replacing $F(t)$ by $F_w(t)=w(t)F(t)$. Alternatively, to reduce the sidelobes in the time signal $\hat{F}(-\dot\omega_0 t)$, one can apply windowing to its Fourier transform, which can be determined as follows:

$$\hat{\hat{F}}(\omega)=\int_{-\infty}^{\infty}\hat{F}(-\dot\omega_0 t)e^{-j\omega t}dt=\int_{-\infty}^{\infty}\hat{F}(\omega')e^{j\omega\omega'/\dot\omega_0}d\omega'/\dot\omega_0=\frac{2\pi}{\dot\omega_0}F(\omega/\dot\omega_0)$$

that is, the time-domain envelope $F(t)$ evaluated at $t=\omega/\dot\omega_0$. Thus, a time window $w(t)$ can just as well be applied in the frequency domain in the form:

$$\hat{\hat{F}}(\omega)=F(\omega/\dot\omega_0)\quad\Rightarrow\quad\hat{\hat{F}}_w(\omega)=w(\omega/\dot\omega_0)F(\omega/\dot\omega_0)$$

Since $w(t)$ is concentrated over $-T/2\le t\le T/2$, the frequency window $w(\omega/\dot\omega_0)$ will be concentrated over

$$-\frac{T}{2}\le\frac{\omega}{\dot\omega_0}\le\frac{T}{2}\quad\Rightarrow\quad-\frac{\Omega}{2}\le\omega\le\frac{\Omega}{2}$$

where $\Omega=\dot\omega_0 T=2\pi B$. For example, a Hamming window, which affords a suppression of the sidelobes by 40 dB, can be applied in the time or frequency domain:

$$\begin{aligned} w(t)&=1+2\alpha\cos\left(\frac{2\pi t}{T}\right),\quad -\frac{T}{2}\le t\le\frac{T}{2}\\ w(\omega/\dot\omega_0)&=1+2\alpha\cos\left(\frac{2\pi\omega}{\Omega}\right),\quad -\frac{\Omega}{2}\le\omega\le\frac{\Omega}{2}\end{aligned} \tag{3.10.23}$$

where $2\alpha = 0.46/0.54$, or, $\alpha = 0.4259$.[†] The time-domain window can be implemented in a straightforward fashion using delays. Writing $w(t)$ in exponential form, we have

$$w(t)=1+\alpha\big[e^{2\pi jt/T}+e^{-2\pi jt/T}\big]$$

The spectrum of $F_w(t)=w(t)F(t)=\big(1+\alpha\big[e^{2\pi jt/T}+e^{-2\pi jt/T}\big]\big)F(t)$ will be:

$$\hat{F}_w(\omega)=\hat{F}(\omega)+\alpha\big[\hat{F}(\omega-2\pi/T)+\hat{F}(\omega+2\pi/T)\big]$$

Thus, the envelope of the compressed signal will be:

$$\begin{aligned}\hat{F}_w(-\dot\omega_0 t)&=\hat{F}(-\dot\omega_0 t)+\alpha\big[\hat{F}(-\dot\omega_0 t-2\pi/T)+\hat{F}(-\dot\omega_0 t+2\pi/T)\big]\\ &=\hat{F}(-\dot\omega_0 t)+\alpha\big[\hat{F}\big(-\dot\omega_0(t+T_{\text{compr}})\big)+\hat{F}\big(-\dot\omega_0(t-T_{\text{compr}})\big)\big]\end{aligned}$$

[†] This definition of $w(t)$ differs from the ordinary Hamming window by a factor of 0.54.

where $T_{\text{compr}} = 2\pi T/\dot{\omega}_0 = 1/B$. It follows that the compressed output will be:

$$E_{\text{compr}}(t) = \sqrt{jBT}\, e^{j\omega_0 t - j\dot{\omega}_0 t^2/2}\, \big[\text{sinc}(Bt) + \alpha\, \text{sinc}(Bt+1) + \alpha\, \text{sinc}(Bt-1)\big] \tag{3.10.24}$$

where $\text{sinc}(x) = \sin(\pi x)/\pi x$, and we wrote $B(t \pm T_{\text{compr}}) = (Bt \pm 1)$. Fig. 3.10.5 shows the Hamming windowed chirped pulse and the corresponding compressed output computed from Eq. (3.10.24).

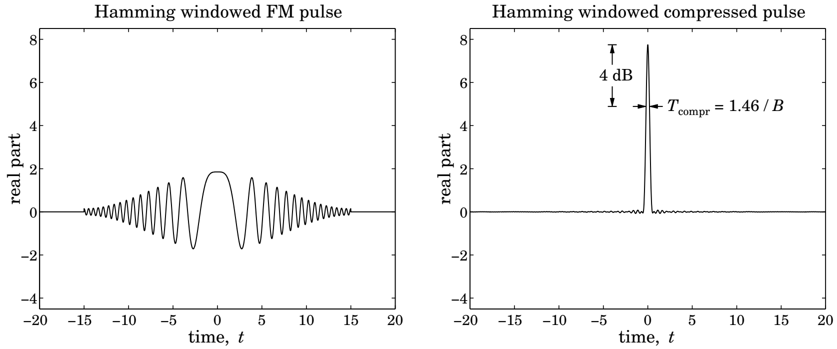


Fig. 3.10.5 Hamming windowed FM pulse and its compressed version, with $T = 30$, $B = 4$.

The price to pay for reducing the sidelobes is a somewhat wider mainlobe width. Measured at the 4-dB level, the width of the compressed pulse is $T_{\text{compr}} = 1.46/B$, as compared with $1/B$ in the unwindowed case.

Matched Filter

A more appropriate choice for the compression filter is the *matched filter*, which maximizes the receiver's SNR. Without getting into the theoretical justification, a filter matched to a transmitted waveform $E(t)$ has the conjugate-reflected impulse response $h(t) = E^*(-t)$ and corresponding frequency response $H(\omega) = \hat{E}^*(\omega)$. In particular for the rectangular chirped pulse, we have:

$$E(t) = \text{rect}\Big(\frac{t}{T}\Big)\, e^{j\omega_0 t + j\dot{\omega}_0 t^2/2} \quad \Rightarrow \quad h(t) = E^*(-t) = \text{rect}\Big(\frac{t}{T}\Big)\, e^{j\omega_0 t - j\dot{\omega}_0 t^2/2} \tag{3.10.25}$$

which differs from our simplified compression filter by the factor $\text{rect}(t/T)$. Its frequency response is given by the conjugate of Eq. (3.10.19)

$$H(\omega) = \sqrt{-\frac{2\pi j}{\dot{\omega}_0}}\, e^{j(\omega-\omega_0)^2/2\dot{\omega}_0}\, D(\omega)\,, \qquad D(\omega) = \frac{\mathcal{F}(w_+) - \mathcal{F}(w_-)}{1-j} \tag{3.10.26}$$

We have seen that the factor $D(\omega)$ is essentially unity within the band $|f - f_0| \le B/2$. Thus again, the matched filter resembles the filter (3.10.10) within this band. The resulting output of the matched filter is remarkably similar to that of Eq. (3.10.18):

$$E_{\text{compr}}(t) = e^{j\omega_0 t}\, T\, \frac{\sin(\pi B|t| - \pi B t^2/T)}{\pi B|t|}\,, \quad \text{for} \quad -T \le t \le T \tag{3.10.27}$$

while it vanishes for $|t| > T$.

In practice, the matched/compression filters are conveniently realized either digitally using digital signal processing (DSP) techniques or using surface acoustic wave (SAW) devices [345]. Similarly, the waveform generator of the chirped pulse may be realized using DSP or SAW methods. A convenient generation method is to send an impulse (or, a broadband pulse) to the input of a filter that has as frequency response $H(\omega) = \hat{E}(\omega)$, so that the impulse response of the filter is the signal $E(t)$ that we wish to generate.

Signal design in radar is a subject in itself and the present discussion was only meant to be an introduction motivated by the similarity to dispersion compensation.

3.11 Further Reading

The topics discussed in this chapter are vast and diverse. The few references given below are inevitably incomplete.

References [130–153] discuss the relationship between group velocity and energy velocity for lossless or lossy media, as well as the issue of electromagnetic field momentum and radiation pressure.

Some references on pulse propagation, spreading, chirping, and dispersion compensation in optical fibers, plasmas, and other media are [154–206], while precursors are discussed in Sommerfeld [1111], Brillouin [154], and [207–219].

Some theoretical and experimental references on fast and negative group velocity are [220–275]. Circuit realizations of negative group delays are discussed in [276–280]. References [281–312] discuss slow light and electromagnetically induced transparency and related experiments.

Some references on chirp radar and pulse compression are [313–352]. These include phase-coding methods, as well as alternative phase modulation methods for Doppler-resistant applications.

3.12 Problems

3.1 Using the definitions (3.2.5), show that the group and phase velocities are related by:

$$v_g = v_p + \beta\,\frac{dv_p}{d\beta}$$

3.2 It was mentioned earlier that when $v_g > c$, the peak of a pulse shifts forward in time as it propagates. With reference to Fig. 3.2.2, let t_{peak} be the time of the peak of the initial pulse $E(0,t)$. First, show that the peak of the propagated pulse $E(z,t)$ occurs at time $t_{\text{prop}} = t_{\text{peak}} + z/v_g$. Then, show that the peak value $E(z, t_{\text{prop}})$ does not depend on the initial peak $E(0, t_{\text{peak}})$ but rather it depends causally on the values $E(0,t)$, for $t_0 \le t \le t_{\text{peak}} - \Delta t$, where $\Delta t = z/c - z/v_g$, which is positive if $v_g > c$. What happens if $0 < v_g < c$ and if $v_g < 0$?

3.3 Consider case 6 of the exactly solvable examples of Sec. 3.3 describing a lossy transmission line with distributed parameters L', C', R', G'. The voltage and current along the line satisfy the so-called telegrapher's equations:

$$\frac{\partial V}{\partial z} + L'\frac{\partial I}{\partial t} + R'I = 0\,, \qquad \frac{\partial I}{\partial z} + C'\frac{\partial V}{\partial t} + G'V = 0 \tag{3.12.1}$$

The voltage impulse response $V(z,t)$ of the line is given by Eq. (3.3.1), where $t_f = z/c$, $a+b = R'/L'$, $a-b = G'/C'$, and $c = 1/\sqrt{L'C'}$:

$$V(z,t) = \delta(t-t_f)e^{-at_f} + e^{-at}\,\frac{I_1\Big(b\sqrt{t^2-t_f^2}\Big)}{\sqrt{t^2-t_f^2}}\,bt_f\,u(t-t_f)$$

Show that the corresponding current $I(z,t)$ is given by

$$\sqrt{\frac{L'}{C'}}\,I(z,t) = \delta(t-t_f)e^{-at_f} + e^{-at}\left[\frac{I_1\Big(b\sqrt{t^2-t_f^2}\Big)}{\sqrt{t^2-t_f^2}}\,bt - bI_0\Big(b\sqrt{t^2-t_f^2}\Big)\right]u(t-t_f)$$

by verifying that V and I satisfy Eqs. (3.12.1). *Hint:* Use the relationships: $I_0'(x) = I_1(x)$ and $I_1'(x) = I_0(x) - I_1(x)/x$ between the Bessel functions $I_0(x)$ and $I_1(x)$.

Next, show that the Fourier transforms of $V(z,t)$ and $I(z,t)$ are:

$$\hat{V}(z,\omega) = e^{-\gamma z}\,, \qquad \hat{I}(z,\omega) = \frac{e^{-\gamma z}}{Z}$$

where γ, Z are the propagation constant and characteristic impedance (see Sec. 10.6):

$$\gamma = jk = \sqrt{R' + j\omega L'}\sqrt{G' + j\omega C'}\,, \qquad Z = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}}$$

3.4 *Computer Experiment—Transient Behavior.* Reproduce the results and graphs of the Figures 3.4.1, 3.4.2, and 3.4.3.

3.5 Consider the propagated envelope of a pulse under the linear approximation of Eq. (3.5.13), that is, $F(z,t) = F(0, t - k_0'z)$, for the case of a complex-valued wavenumber, $k_0' = \beta_0' - j\alpha_0'$. For a gaussian envelope:

$$F(z,t) = F(0, t-k_0'z) = \exp\left[-\frac{\big(t - (\beta_0' - j\alpha_0')z\big)^2}{2\tau_0^2}\right]$$

Determine an expression for its magnitude $|F(z,t)|$. Then show that the maximum of $|F(z,t)|$ with respect to t at a given fixed z is moving with the group velocity $v_g = 1/\beta_0'$.

Alternatively, at fixed t show that the maximum with respect to z of the snapshot $|F(z,t)|$ is moving with velocity [160]:

$$v = \frac{\beta_0'}{\beta_0'^2 - \alpha_0'^2}$$

3.6 Consider the propagating wave $E(z,t) = F(z,t)e^{j\omega_0 t - jk_0 z}$. Assuming the quadratic approximation (3.5.9), show that the envelope $F(z,t)$ satisfies the partial differential equation:

$$\left(\frac{\partial}{\partial z} + k_0'\frac{\partial}{\partial t} - j\frac{k_0''}{2}\frac{\partial^2}{\partial t^2}\right)F(z,t) = 0$$

Show that the envelope impulse response $g(z,t)$ of Eq. (3.5.16) also satisfies this equation. And that so does the gaussian pulse of Eq. (3.6.4).

3.7 Let $F(z,t)$ be the narrowband envelope of a propagating pulse as in Eq. (3.5.5). Let $z(t)$ be a point on the snapshot $F(z,t)$ that corresponds to a particular constant value of the envelope, that is, $F(z(t),t) =$ constant. Show that the point $z(t)$ is moving with velocity:

$$\dot{z}(t) = -\frac{\partial_t F}{\partial_z F}$$

Under the linear approximation of Eq. (3.5.13), show that the above expression leads to the group velocity $\dot{z}(t) = 1/k_0'$.

Alteratively, use the condition $|F(z,t)|^2 =$ constant, and show that in this case

$$\dot{z}(t) = -\frac{\mathrm{Re}(\partial_t F/F)}{\mathrm{Re}(\partial_z F/F)}$$

Under the linear approximation and assuming that the initial envelope $F(0,t)$ is real-valued, show that $\dot{z} = 1/\,\mathrm{Re}(k_0')$.

3.8 Given the narrowband envelope $F(z,t)$ of a propagating pulse as in Eq. (3.5.5), show that it satisfies the identity:

$$e^{-j(k-k_0)z}\hat{F}(0, \omega - \omega_0) = \int_{-\infty}^{\infty} F(z,t)e^{-j(\omega-\omega_0)t}\,dt$$

Define the "centroid" time $t(z)$ by the equation

$$t(z) = \frac{\int_{-\infty}^{\infty} t\,F(z,t)\,dt}{\int_{-\infty}^{\infty} F(z,t)\,dt}$$

Using the above identity, show that $t(z)$ satisfies the equation:

$$t(z) = t(0) + k_0'z \tag{3.12.2}$$

Therefore, $t(z)$ may be thought of as a sort of group delay. Note that no approximations are needed to obtain Eq. (3.12.2).

3.9 Consider the narrowband envelope $F(z,t)$ of a propagating pulse $E(z,t) = F(z,t)e^{j\omega_0 t - jk_0 z}$ and assume that the medium is lossless so that $k(\omega)$ is real-valued. Show the identity

$$\int_{-\infty}^{\infty} |E(z,t)|^2 e^{-j\omega t}\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{jk(\omega')z}\,\hat{E}^*(0,\omega')\,e^{-jk(\omega'+\omega)z}\,\hat{E}(0,\omega'+\omega)\,d\omega'$$

Define the average time delay and average inverse group velocity through:

$$\bar{t}(z) = \frac{\int_{-\infty}^{\infty} t\,|F(z,t)|^2\,dt}{\int_{-\infty}^{\infty} |F(z,t)|^2\,dt}\,, \qquad \bar{k}_0' = \frac{\int_{-\infty}^{\infty} k'(\omega_0+\omega)\,|\hat{F}(0,\omega)|^2\,d\omega}{\int_{-\infty}^{\infty} |\hat{F}(0,\omega)|^2\,d\omega}$$

where $\hat{F}(0,\omega)$ is defined in Eq. (3.5.2). Using the above identity, show the relationship:

$$\bar{t}(z) = \bar{t}(0) + \bar{k}_0'z$$

3.10 *Computer Experiment—Propagation with Negative Group Velocity.* Consider the pulse propagation experiment described in Figs. 3.9.3 and 3.9.4, which is a variation of the experiment in Ref. [234]. The wavenumbers in vacuum, in the absorption and gain media will be denoted by k_v, k_a, k_g. They can be calculated from Eqs. (3.9.1)–(3.9.6) with $f = 0, +1, -1$, respectively.

Let $E(t)$ and $\hat{E}(\omega)$ be the initially launched waveform and its Fourier transform on the vacuum side of the interface with the absorbing medium at $z = 0$. Because the refractive indices $n(\omega)$ are very nearly unity, we will ignore all the reflected waves and assume that the wave enters the successive media with unity transmission coefficient.

a. Show that the wave will be given as follows in the successive media shown in Fig. 3.9.3:

$$E(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{E}(\omega)d\omega\begin{cases} e^{j\omega t-jk_v z} & \text{if } z\le 0\\ e^{j\omega t-jk_a z} & \text{if } 0<z\le a\\ e^{j\omega t-jk_a a-jk_v(z-a)} & \text{if } a<z\le 3a\\ e^{j\omega t-j(k_a+2k_v)a-jk_g(z-3a)} & \text{if } 3a<z\le 4a\\ e^{j\omega t-j(k_a+2k_v+k_g)a-jk_v(z-4a)} & \text{if } 4a<z \end{cases} \tag{3.12.3}$$

Thus, in each region, the pulse will have the following form, with appropriate definitions of the wavenumbers $q(\omega)$, $k(\omega)$, and offset d:

$$E(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{E}(\omega)e^{j\omega t-jqa-jk(z-d)}d\omega \tag{3.12.4}$$

b. Consider, next, a gaussian pulse with width τ_0, modulating a carrier ω_0, defined at $z=0$ as follows:

$$E(t)=e^{-t^2/2\tau_0^2}e^{j\omega_0 t}\quad\Leftrightarrow\quad \hat{E}(\omega)=\sqrt{2\pi\tau_0^2}\,e^{-\tau_0^2(\omega-\omega_0)^2/2} \tag{3.12.5}$$

Assuming a sufficiently narrow bandwidth (small $1/\tau_0$ or large τ_0,) the wavenumbers $q(\omega)$ and $k(\omega)$ in Eq. (3.12.4) can be expanded up to second order about the carrier frequency ω_0 giving:

$$\begin{aligned} q(\omega) &= q_0+q_0'(\omega-\omega_0)+q_0''(\omega-\omega_0)^2/2\\ k(\omega) &= k_0+k_0'(\omega-\omega_0)+k_0''(\omega-\omega_0)^2/2 \end{aligned} \tag{3.12.6}$$

where the quantities $q_0=q(\omega_0)$, $q_0'=q'(\omega_0)$, etc., can be calculated from Eqs. (3.9.1)–(3.9.6). Inserting these expansions into Eq. (3.12.4), show that the pulse waveform is given by:

$$\begin{aligned} E(z,t) &= e^{j\omega_0 t-jq_0 a-jk_0(z-d)}\sqrt{\frac{\tau_0^2}{\tau_0^2+jq_0''a+jk_0''(z-d)}}\\ &\quad\cdot\exp\left[-\frac{\left(t-q_0'a-k_0'(z-d)\right)^2}{2\left(\tau_0^2+jq_0''a+jk_0''(z-d)\right)}\right] \end{aligned} \tag{3.12.7}$$

c. Assume the following values of the various parameters:

$$c=1,\quad \omega_p=1,\quad \gamma=0.01,\quad \omega_r=5,\quad \tau_0=40,\quad a=50,\quad \omega_0=5.35$$

The carrier frequency ω_0 is chosen to lie in the right wing of the resonance and lies in the negative-group-velocity range for the gain medium (this range is approximately $[5.005,\ 5.5005]$ in the above frequency units.)

Calculate the values of the parameters q_0, q_0', q_0'', k_0, k_0', k_0'' within the various ranges of z as defined by Eq. (3.12.3), and present these values in a table form.

Thus, $E(z,t)$ can be evaluated for each value of t and for all the z's in the four ranges. Eq. (3.12.7) can easily be vectorized for each scalar t and a vector of z's.

Make a MATLAB movie of the pulse envelope $|E(z,t)|$, that is, for each successive t, plot the envelope versus z. Take z to vary over $-2a\le z\le 6a$ and t over $-100\le t\le 300$. Such a movie can be made with the following code fragment:

```
z = -2*a : 6*a;        % vector of z's
t = -100 : 300;        % vector of t's

for i=1:length(t),

  % here insert code that calculates the vector E = E(z, t(i))

  plot(z/a, abs(E));                            % plot as function of z
  xlim([-2,6]); xtick([0,1,3,4]); grid          % keep axes the same
  ylim([0,1]); ytick(0:1:1);                    % xtick, ytick are part of ewa
  text(-1.8, 0.35, strcat('t=',num2str(t(i))), 'fontsize', 15);

  F(:,i) = getframe;          % save current frame

end

movie(F);     % replay movie - check syntax of movie() for playing options
```

Discuss your observations, and explain what happens within the absorption and gain media. An example of such a movie may be seen by running the file `grvmovie1.m` in the `movies` subdirectory of the ewa toolbox.

d. Reproduce the graphs of Fig. 3.9.4 by evaluating the snapshots at the time instants:

$$t=[-50,\ 0,\ 40,\ 120,\ 180,\ 220,\ 230,\ 240,\ 250,\ 260]$$

e. For both the absorbing and the gain media, plot the real and imaginary parts of the refractive index $n=n_r-jn_i$ and the real part of the group index n_g versus frequency in the interval $4\le\omega\le 6$. Indicate on the graph the operating frequency points. For the gain case, indicate the ranges over which $\mathrm{Re}(n_g)$ is negative.

f. Repeat Parts c–d for the carrier frequency $\omega_0=5.8$ which lies in the superluminal range $0<\mathrm{Re}(n_g)<1$.

3.11 Consider Eqs. (3.9.1)–(3.9.6) for the single-resonance Lorentz model that was used in the previous experiment. Following [234], define the detuning parameters:

$$\xi=\frac{\omega-\omega_r}{\omega_p},\quad \xi_0=\frac{\omega_0-\omega_r}{\omega_p} \tag{3.12.8}$$

and make the following assumptions regarding the range of these quantities:

$$\gamma\ll\omega_p\ll\omega_r\quad\text{and}\quad\frac{\omega_p}{\omega_r}\ll|\xi|\ll\frac{\omega_r}{\omega_p} \tag{3.12.9}$$

Thus, $\gamma/\omega_p\ll 1$, $\omega_p/\omega_r\ll 1$, and ξ,ξ_0 can be taken to be order of 1. In the above experiment, they were $\xi_0=0.35$ and $\xi_0=0.8$.

Show that the wavenumber $k(\omega)$, and its first and second derivatives $k'(\omega)$, $k''(\omega)$, can be expressed approximately to first order in the quantities γ/ω_p and ω_p/ω_r as follows [234]:

$$\begin{aligned} k(\omega) &= \frac{\omega}{c}\left[1-\frac{f}{4\xi}\frac{\omega_p}{\omega_r}-j\frac{f}{8\xi^2}\left(\frac{\omega_p}{\omega_r}\right)\left(\frac{\gamma}{\omega_p}\right)\right]\\ k'(\omega) &= \frac{1}{c}\left[1+\frac{f}{4\xi^2}+j\frac{f}{4\xi^3}\left(\frac{\gamma}{\omega_p}\right)\right]\\ k''(\omega) &= -\frac{1}{c\omega_p}\left[\frac{f}{2\xi^3}+j\frac{3f}{4\xi^4}\left(\frac{\gamma}{\omega_p}\right)\right] \end{aligned}$$

The group velocity v_g is obtained from the real part of $k'(\omega)$:

$$\frac{1}{v_g} = \mathrm{Re}[k'(\omega)] = \frac{1}{c}\left[1 + \frac{f}{4\xi^2}\right] \quad \Rightarrow \quad v_g = \frac{c}{1 + \dfrac{f}{4\xi^2}}$$

Thus, $v_g = c$ in vacuum $(f = 0)$ and $v_g < c$ in the absorbing medium $(f = 1)$. For the gain medium $(f = -1)$, we have $v_g < 0$ if $|\xi| < 1/2$, and $v_g > c$ if $|\xi| > 1/2$.

Verify that this approximation is adequate for the numerical values given in the previous problem.

3.12 Consider a chirped pulse whose spectrum has an ideal rectangular shape and an ideal quadratic phase, where $\Omega = 2\pi B$ and $\dot\omega_0 = 2\pi B/T$:

$$\hat{E}(\omega) = \sqrt{\frac{2\pi j}{\dot\omega_0}}\,\mathrm{rect}\left(\frac{\omega - \omega_0}{\Omega}\right) e^{-j(\omega-\omega_0)^2/2\dot\omega_0}$$

This is the ideal spectrum that all waveforms in chirp radar strive to have. Show that the corresponding time signal $E(t)$ is given in terms of the Fresnel function $\mathcal{F}(x)$ by

$$E(t) = F(t)\, e^{j\omega_0 t + j\dot\omega_0 t^2/2}, \quad F(t) = \frac{\mathcal{F}(\tau_+) - \mathcal{F}(\tau_-)}{1 - j}, \quad \tau_\pm = \sqrt{2BT}\left(\pm\frac{1}{2} - \frac{t}{T}\right)$$

Show that the output of the compression filter (3.10.10) is given by

$$E_{\mathrm{compr}}(t) = \sqrt{jBT}\,\frac{\sin(\pi B t)}{\pi B t}\, e^{j\omega_0 t}$$

3.13 *Computer Experiment—Pulse Compression.* Take $T = 30$, $B = 4$, $f_0 = 0$. Plot the real parts of the signals $E(t)$ and $E_{\mathrm{compr}}(t)$ of the previous problem versus t over the interval $-T \le t \le T$. Some example graphs are shown in Fig. 3.12.1.

3.14 Consider the following chirped pulse, where $\dot\omega_0 = 2\pi B/T$:

$$E(t) = \frac{\sin(\pi t/T)}{\pi t/T}\, e^{j\omega_0 t + j\dot\omega_0 t^2/2}$$

Show that the output of the compression filter (3.10.10) is given by

$$E_{\mathrm{compr}}(t) = \sqrt{jBT}\, e^{j\omega_0 t - j\dot\omega_0 t^2/2}\,\mathrm{rect}(Bt)$$

Moreover, show that the spectrum of $E(t)$ is given in terms of the Fresnel function $\mathcal{F}(x)$ as follows:

$$\hat{E}(\omega) = \frac{\mathcal{F}(w_+) - \mathcal{F}(w_-)}{1 - j}, \quad w_\pm = \sqrt{\frac{2}{BT}}\left(\pm\frac{1}{2} - (f - f_0)T\right)$$

3.15 *Computer Experiment—Pulse Compression.* Take $T = 30$, $B = 4$, $f_0 = 0$.

a. Plot the real parts of the signals $E(t)$ and $E_{\mathrm{compr}}(t)$ of the previous problem over the interval $-T \le t \le T$. Some example graphs are shown in Fig. 3.12.2.

b. Plot the magnitude spectrum $|\hat{E}(\omega)|$ in dB versus frequency over the interval $-B/2 \le f \le B/2$ (normalize the spectrum to its maximum at $f = f_0$.) Verify that the spectrum lies essentially within the desired bandwidth B and determine its 4-dB width.

c. Write $\hat{E}(\omega)$ in the form:

$$\hat{E}(\omega) = e^{-j(\omega-\omega_0)^2/2\dot\omega_0}\, D(\omega), \quad D(\omega) = \frac{\mathcal{F}(w_+) - \mathcal{F}(w_-)}{1 - j}\, e^{j(\omega-\omega_0)^2/2\dot\omega_0}$$

Plot the residual phase spectrum $\mathrm{Arg}[D(\omega)]$ over the above frequency interval. Verify that it remains essentially flat, confirming that the phase of $\hat{E}(\omega)$ has the expected quadratic dependence on ω. Show that the small residual constant phase is numerically is equal to the phase of the complex number $(1 + j)\mathcal{F}(1/\sqrt{2BT})$.

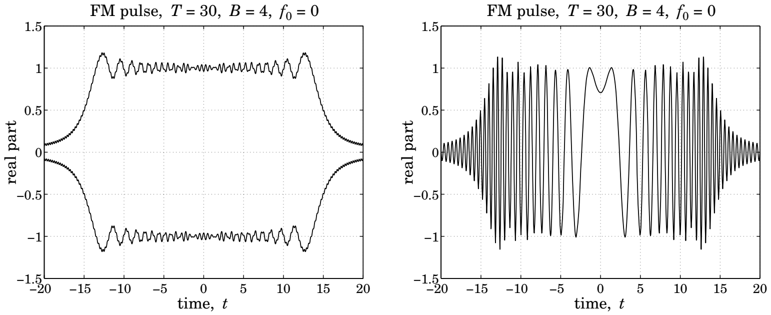


Fig. 3.12.1 Example graphs for Problem 3.13.

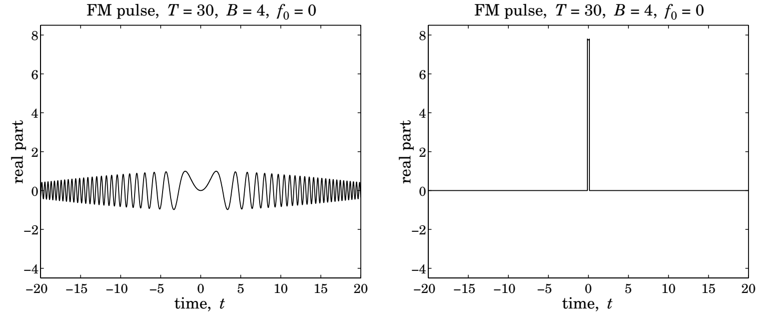


Fig. 3.12.2 Example graphs for Problem 3.15.

3.16 Consider the chirped gaussian pulse of effective duration T, where $\dot\omega_0 = \Omega/T$:

$$E(t) = e^{-t^2/2T^2}\, e^{j\omega_0 t + j\dot\omega_0 t^2/2}, \quad -\infty < t < \infty$$

Show that the output of the compression filter (3.10.10) is given by

$$E_{\mathrm{compr}}(t) = \sqrt{j\Omega T}\, e^{j\omega_0 t - j\dot\omega_0 t^2/2}\, e^{-\Omega^2 t^2/2}$$

which has an effective duration of $1/\Omega$. Show that the spectrum of $E(t)$ is given by:

$$\hat{E}(\omega) = \sqrt{\frac{2\pi j}{\dot\omega_0}}\, e^{-j(\omega-\omega_0)^2/2/\dot\omega_0}\sqrt{\frac{\Omega T}{\Omega T + j}}\,\exp\left[-\frac{(\omega-\omega_0)^2}{2\dot\omega_0(\Omega T + j)}\right]$$

Show that in the limit of large time-bandwidth product, $\Omega T \gg 1$, the last exponential factor becomes

$$\exp\left[-\frac{(\omega-\omega_0)^2}{2\Omega^2}\right]$$

which shows that the effective width of the chirped spectrum is Ω.

3.17 *Stationary-Phase Approximation.* Consider a radar waveform $E(t)=F(t)e^{j\theta(t)}$, with envelope $F(t)$ and phase $\theta(t)$.

a. Using the stationary-phase approximation of Eq. (F.22) of Appendix F, show that the spectrum of $E(t)$ can be expressed approximately as:

$$\hat{E}(\omega)=\int_{-\infty}^{\infty}F(t)e^{j\theta(t)}e^{-j\omega t}\,dt\simeq\sqrt{\frac{2\pi j}{\ddot{\theta}(t_\omega)}}\,E(t_\omega)e^{-j\omega t_\omega}=\sqrt{\frac{2\pi j}{\ddot{\theta}(t_\omega)}}\,F(t_\omega)e^{j\theta(t_\omega)}e^{-j\omega t_\omega}$$

where t_ω is the solution of the equation $\dot{\theta}(t)=\omega$, obtained by applying the stationary-phase approximation to the phase function $\phi(t)=\theta(t)-\omega t$.

b. For the case of a linearly chirped signal $E(t)=F(t)e^{j\omega_0 t+j\dot{\omega}_0 t^2/2}$, show that the above approximation reads:

$$\hat{E}(\omega)\simeq\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,F\left(\frac{\omega-\omega_0}{\dot{\omega}_0}\right)e^{-j(\omega-\omega_0)^2/2\dot{\omega}_0}$$

Thus, it has the usual quadratic phase dispersion. Show that if $F(t)$ has finite duration over the time interval $|t|\le T/2$, then, the above approximate spectrum is sharply confined within the band $|\omega-\omega_0|\le\Omega/2$, with bandwidth $\Omega=\dot{\omega}_0 T$.

c. Consider the inverse Fourier transform of the above expression:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\sqrt{\frac{2\pi j}{\dot{\omega}_0}}\,F\left(\frac{\omega-\omega_0}{\dot{\omega}_0}\right)e^{-j(\omega-\omega_0)^2/2\dot{\omega}_0}e^{j\omega t}d\omega$$

Define the phase function $\phi(\omega)=\omega t-(\omega-\omega_0)^2/2\dot{\omega}_0$. By applying the stationary-phase approximation to the above integral with respect to the phase function $\phi(\omega)$, show that the above inverse Fourier transform is precisely equal to the original chirped signal, that is, $E(t)=F(t)e^{j\omega_0 t+j\dot{\omega}_0 t^2/2}$.

d. Apply the compression filter (3.10.10) to the approximate spectrum of Part b, and show that the corresponding compressed signal is in the time domain:

$$E_{\text{compr}}(t)=\sqrt{\frac{j\dot{\omega}_0}{2\pi}}\,e^{j\omega_0 t}\,\hat{F}(-\dot{\omega}_0 t)$$

where $\hat{F}(\omega)$ is the Fourier transform of $F(t)$. This is similar, but not quite identical, to the exact expression (3.10.14).

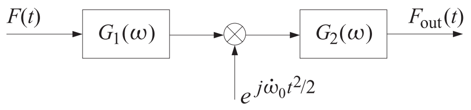
e. Show that Part c is a general result. Consider the stationary-phase approximation spectrum of Part a. Its inverse Fourier transform is:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\sqrt{\frac{2\pi j}{\ddot{\theta}(t_\omega)}}\,F(t_\omega)e^{j\theta(t_\omega)}e^{-j\omega t_\omega}e^{j\omega t}d\omega$$

Define the phase function $\phi(\omega)=\theta(t_\omega)-\omega t_\omega+\omega t$. Show that the stationary-phase approximation applied to this integral with respect to the phase function $\phi(\omega)$ recovers the original waveform $E(t)=F(t)e^{j\theta(t)}$.

[*Hint:* the condition $\dot{\theta}(t_\omega)=\omega$ implies $\ddot{\theta}(t_\omega)\,(dt_\omega/d\omega)=1$.]

3.18 An envelope signal $F(t)$ is processed through two successive pulse compression filters with chirping parameters $\dot{\omega}_1$ and $\dot{\omega}_2$, as shown below.



where $G_i(\omega)=e^{j\omega^2/2\dot{\omega}_i}$, $i=1,2$. Show that if the chirping parameter of the intermediate quadratic modulation is chosen to be $\dot{\omega}_0=\dot{\omega}_1+\dot{\omega}_2$, then the overall output is a time-scaled version of the input:

$$F_{\text{out}}(t)=j\sqrt{\frac{\dot{\omega}_2}{\dot{\omega}_1}}\,e^{j\dot{\omega}_0\dot{\omega}_2 t^2/2\dot{\omega}_1}\,F\left(-\frac{\dot{\omega}_2 t}{\dot{\omega}_1}\right)$$

4

Propagation in Birefringent Media

4.1 Linear and Circular Birefringence

In this chapter, we discuss wave propagation in anisotropic media that are linearly or circularly *birefringent*. In such media, uniform plane waves can be decomposed in two *orthogonal* polarization states (linear or circular) that propagate with two different speeds. The two states develop a *phase difference* as they propagate, which alters the total polarization of the wave. Such media are used in the construction of devices for generating different polarizations.

Linearly birefringent materials can be used to change one polarization into another, such as changing linear into circular. Examples are the so-called *uniaxial* crystals, such as calcite, quartz, ice, tourmaline, and sapphire.

Optically active or *chiral* media are circularly birefringent. Examples are sugar solutions, proteins, lipids, nucleic acids, amino acids, DNA, vitamins, hormones, and virtually most other natural substances. In such media, circularly polarized waves go through unchanged, with left- and right-circular polarizations propagating at different speeds. This difference causes linearly polarized waves to have their polarization plane rotate as they propagate—an effect known as *natural optical rotation*.

A similar but not identical effect—the *Faraday rotation*—takes place in *gyroelectric* media, which are ordinary isotropic materials (glass, water, conductors, plasmas) subjected to constant external magnetic fields that break their isotropy. *Gyromagnetic* media, such as ferrites subjected to magnetic fields, also become circularly birefringent.

We discuss all four birefringent cases (linear, chiral, gyroelectric, and gyromagnetic) and the type of constitutive relationships that lead to the corresponding birefringent behavior. We begin by casting Maxwell's equations in different polarization bases.

An arbitrary polarization can be expressed uniquely as a linear combination of two polarizations along two orthogonal directions.[†] For waves propagating in the z-direction, we may use the two *linear* directions $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}\}$, or the two *circular* ones for right and left polarizations $\{\hat{\mathbf{e}}_+, \hat{\mathbf{e}}_-\}$, where $\hat{\mathbf{e}}_+ = \hat{\mathbf{x}} - j\hat{\mathbf{y}}$ and $\hat{\mathbf{e}}_- = \hat{\mathbf{x}} + j\hat{\mathbf{y}}$.[‡] Indeed, we have the following identity relating the linear and circular bases:

[†] For complex-valued vectors $\mathbf{e}_1, \mathbf{e}_2$, orthogonality is defined with conjugation: $\mathbf{e}_1^* \cdot \mathbf{e}_2 = 0$.

[‡] Note that $\hat{\mathbf{e}}_\pm$ satisfy: $\hat{\mathbf{e}}_\pm^* \cdot \hat{\mathbf{e}}_\pm = 2$, $\hat{\mathbf{e}}_+^* \cdot \hat{\mathbf{e}}_- = 0$, $\hat{\mathbf{e}}_+ \times \hat{\mathbf{e}}_- = 2j\,\hat{\mathbf{z}}$, and $\hat{\mathbf{z}} \times \hat{\mathbf{e}}_\pm = \pm j\,\hat{\mathbf{e}}_\pm$.

$$\boxed{\mathbf{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y = \hat{\mathbf{e}}_+E_+ + \hat{\mathbf{e}}_-E_-}\,, \quad \text{where} \quad E_\pm = \frac{1}{2}(E_x \pm jE_y) \tag{4.1.1}$$

The circular components E_+ and E_- represent right and left polarizations (in the IEEE convention) if the wave is moving in the positive z-direction, but left and right if it is moving in the negative z-direction.

Because the propagation medium is not isotropic, we need to start with the source-free Maxwell's equations before we assume any particular constitutive relationships:

$$\nabla\times\mathbf{E} = -j\omega\mathbf{B}\,, \quad \nabla\times\mathbf{H} = j\omega\mathbf{D}\,, \quad \nabla\cdot\mathbf{D} = 0\,, \quad \nabla\cdot\mathbf{B} = 0 \tag{4.1.2}$$

For a uniform plane wave propagating in the z-direction, we may replace the gradient by $\nabla = \hat{\mathbf{z}}\,\partial_z$. It follows that the curls $\nabla\times\mathbf{E} = \hat{\mathbf{z}}\times\partial_z\mathbf{E}$ and $\nabla\times\mathbf{H} = \hat{\mathbf{z}}\times\partial_z\mathbf{H}$ will be transverse to the z-direction. Then, Faraday's and Ampère's laws imply that $D_z = 0$ and $B_z = 0$, and hence both of Gauss' laws are satisfied. Thus, we are left only with:

$$\begin{aligned} \hat{\mathbf{z}}\times\partial_z\mathbf{E} &= -j\omega\mathbf{B} \\ \hat{\mathbf{z}}\times\partial_z\mathbf{H} &= j\omega\mathbf{D} \end{aligned} \tag{4.1.3}$$

These equations do not "see" the components E_z, H_z. However, in all the cases that we consider here, the conditions $D_z = B_z = 0$ will imply also that $E_z = H_z = 0$. Thus, all fields are transverse, for example, $\mathbf{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y = \hat{\mathbf{e}}_+E_+ + \hat{\mathbf{e}}_-E_-$. Equating x, y components in the two sides of Eq. (4.1.3), we find in the linear basis:

$$\boxed{\begin{aligned} \partial_z E_x &= -j\omega B_y\,, & \partial_z E_y &= j\omega B_x \\ \partial_z H_y &= -j\omega D_x\,, & \partial_z H_x &= j\omega D_y \end{aligned}} \quad \text{(linear basis)} \tag{4.1.4}$$

Using the vector property $\hat{\mathbf{z}}\times\hat{\mathbf{e}}_\pm = \pm j\,\hat{\mathbf{e}}_\pm$ and equating circular components, we obtain the circular-basis version of Eq. (4.1.3) (after canceling some factors of j):

$$\boxed{\begin{aligned} \partial_z E_\pm &= \mp\omega B_\pm \\ \partial_z H_\pm &= \pm\omega D_\pm \end{aligned}} \quad \text{(circular basis)} \tag{4.1.5}$$

4.2 Uniaxial and Biaxial Media

In uniaxial and biaxial homogeneous anisotropic dielectrics, the $D{-}E$ constitutive relationships are given by the following diagonal forms, where in the biaxial case all diagonal elements of the permittivity matrix are distinct:

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_e & 0 & 0 \\ 0 & \epsilon_o & 0 \\ 0 & 0 & \epsilon_o \end{bmatrix}\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}\begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{4.2.1}$$

For the uniaxial case, the x-axis is taken to be the *extraordinary* axis with $\epsilon_1 = \epsilon_e$, whereas the y and z axes are *ordinary* axes with permittivities $\epsilon_2 = \epsilon_3 = \epsilon_o$.

The ordinary z-axis was chosen to be the propagation direction in order for the transverse x, y axes to correspond to two different permittivities. In this respect, the uniaxial and biaxial cases are similar, and therefore, we will work with the biaxial case. Setting $D_x = \epsilon_1 E_x$ and $D_y = \epsilon_2 E_y$ in Eq. (4.1.4) and assuming $\mathbf{B} = \mu_0\mathbf{H}$, we have:

$$\begin{aligned} \partial_z E_x &= -j\omega\mu_0 H_y\,, & \partial_z E_y &= j\omega\mu_0 H_x \\ \partial_z H_y &= -j\omega\epsilon_1 E_x\,, & \partial_z H_x &= j\omega\epsilon_2 E_y \end{aligned} \tag{4.2.2}$$

Differentiating these once more with respect to z, we obtain the decoupled Helmholtz equations for the x-polarized and y-polarized components:

$$\begin{aligned} \partial_z^2 E_x &= -\omega^2\mu_0\epsilon_1 E_x \\ \partial_z^2 E_y &= -\omega^2\mu_0\epsilon_2 E_y \end{aligned} \tag{4.2.3}$$

The forward-moving solutions are:

$$\begin{aligned} E_x(z) &= Ae^{-jk_1 z}\,, & k_1 &= \omega\sqrt{\mu_0\epsilon_1} = k_0 n_1 \\ E_y(z) &= Be^{-jk_2 z}\,, & k_2 &= \omega\sqrt{\mu_0\epsilon_2} = k_0 n_2 \end{aligned} \tag{4.2.4}$$

where $k_0 = \omega\sqrt{\mu_0\epsilon_0} = \omega/c_0$ is the free-space wavenumber and we defined the refractive indices $n_1 = \sqrt{\epsilon_1/\epsilon_0}$ and $n_2 = \sqrt{\epsilon_2/\epsilon_0}$. Therefore, the total transverse field at $z = 0$ and at distance $z = l$ inside the medium will be:

$$\begin{aligned} \mathbf{E}(0) &= \hat{\mathbf{x}}A + \hat{\mathbf{y}}B \\ \mathbf{E}(l) &= \hat{\mathbf{x}}Ae^{-jk_1 l} + \hat{\mathbf{y}}Be^{-jk_2 l} = \big[\hat{\mathbf{x}}A + \hat{\mathbf{y}}Be^{j(k_1-k_2)l}\big]e^{-jk_1 l} \end{aligned} \tag{4.2.5}$$

The relative phase $\phi = (k_1 - k_2)l$ between the x- and y-components introduced by the propagation is called *retardance*:

$$\boxed{\phi = (k_1 - k_2)l = (n_1 - n_2)k_0 l = (n_1 - n_2)\frac{2\pi l}{\lambda}} \tag{4.2.6}$$

where λ is the free-space wavelength. Thus, the polarization nature of the field keeps changing as it propagates.

In order to change linear into circular polarization, the wave may be launched into the birefringent medium with a linear polarization having equal x- and y-components. After it propagates a distance l such that $\phi = (n_1 - n_2)k_0 l = \pi/2$, the wave will have changed into left-handed circular polarization:

$$\begin{aligned} \mathbf{E}(0) &= A\,(\hat{\mathbf{x}} + \hat{\mathbf{y}}) \\ \mathbf{E}(l) &= A\,(\hat{\mathbf{x}} + \hat{\mathbf{y}}\,e^{j\phi})\,e^{-jk_1 l} = A\,(\hat{\mathbf{x}} + j\,\hat{\mathbf{y}})\,e^{-jk_1 l} \end{aligned} \tag{4.2.7}$$

Polarization-changing devices that employ this property are called *retarders* and are shown in Fig. 4.2.1. The above example is referred to as a *quarter-wave retarder* because the condition $\phi = \pi/2$ may be written as $(n_1 - n_2)l = \lambda/4$.

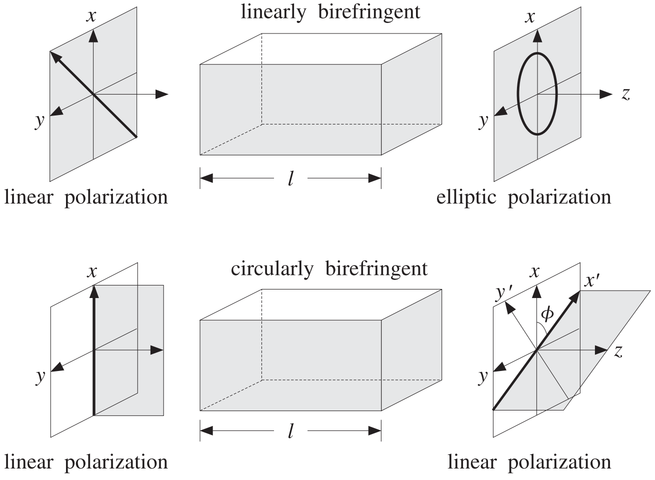


Fig. 4.2.1 Linearly and circularly birefringent retarders.

4.3 Chiral Media

Ever since the first experimental observations of optical activity by Arago and Biot in the early 1800s and Fresnel's explanation that optical rotation is due to circular birefringence, there have been many attempts to explain it at the molecular level. Pasteur was the first to postulate that optical activity is caused by the chirality of molecules.

There exist several versions of constitutive relationships that lead to circular birefringence [679–695]. For single-frequency waves, they are all equivalent to each other. For our purposes, the following so-called Tellegen form is the most convenient [33]:

$$\boxed{\begin{aligned}\mathbf{D} &= \epsilon\mathbf{E} - j\chi\mathbf{H}\\ \mathbf{B} &= \mu\mathbf{H} + j\chi\mathbf{E}\end{aligned}} \qquad \text{(chiral media)} \tag{4.3.1}$$

where χ is a parameter describing the chirality properties of the medium.

It can be shown that the reality (for a lossless medium) and positivity of the energy density function $(\mathbf{E}^* \cdot \mathbf{D} + \mathbf{H}^* \cdot \mathbf{B})/2$ requires that the constitutive matrix

$$\begin{bmatrix} \epsilon & -j\chi \\ j\chi & \mu \end{bmatrix}$$

be hermitian and positive definite. This implies that ϵ, μ, χ are real, and furthermore, that $|\chi| < \sqrt{\mu\epsilon}$. Using Eqs. (4.3.1) in Maxwell's equations (4.1.5), we obtain:

$$\begin{aligned}\partial_z E_\pm &= \mp\omega B_\pm = \mp\omega(\mu H_\pm + j\chi E_\pm)\\ \partial_z H_\pm &= \pm\omega D_\pm = \pm\omega(\epsilon E_\pm - j\chi H_\pm)\end{aligned} \tag{4.3.2}$$

Defining $c = 1/\sqrt{\mu\epsilon}$, $\eta = \sqrt{\mu/\epsilon}$, $k = \omega/c = \omega\sqrt{\mu\epsilon}$, and the following real-valued dimensionless parameter $a = c\chi = \chi/\sqrt{\mu\epsilon}$ (so that $|a| < 1$), we may rewrite Eqs. (4.3.2) in the following matrix forms:

$$\frac{\partial}{\partial z}\begin{bmatrix} E_\pm \\ \eta H_\pm \end{bmatrix} = \mp \begin{bmatrix} jka & k \\ -k & jka \end{bmatrix}\begin{bmatrix} E_\pm \\ \eta H_\pm \end{bmatrix} \tag{4.3.3}$$

These matrix equations may be diagonalized by appropriate linear combinations. For example, we define the right-polarized (forward-moving) and left-polarized (backward-moving) waves for the $\{E_+, H_+\}$ case:

$$\begin{aligned} E_{R+} &= \frac{1}{2}[E_+ - j\eta H_+] \\ E_{L+} &= \frac{1}{2}[E_+ + j\eta H_+] \end{aligned} \quad\Leftrightarrow\quad \begin{aligned} E_+ &= E_{R+} + E_{L+} \\ H_+ &= -\frac{1}{j\eta}[E_{R+} - E_{L+}] \end{aligned} \tag{4.3.4}$$

It then follows from Eq. (4.3.3) that $\{E_{R+}, E_{L+}\}$ will satisfy the decoupled equations:

$$\frac{\partial}{\partial z}\begin{bmatrix} E_{R+} \\ E_{L+} \end{bmatrix} = \begin{bmatrix} -jk_+ & 0 \\ 0 & jk_- \end{bmatrix}\begin{bmatrix} E_{R+} \\ E_{L+} \end{bmatrix} \quad\Rightarrow\quad \begin{aligned} E_{R+}(z) &= A_+ e^{-jk_+ z} \\ E_{L+}(z) &= B_+ e^{jk_- z} \end{aligned} \tag{4.3.5}$$

where k_+, k_- are defined as follows:

$$\boxed{k_\pm = k(1 \pm a) = \omega\left(\sqrt{\mu\epsilon} \pm \chi\right)} \tag{4.3.6}$$

We may also define circular refractive indices by $n_\pm = k_\pm/k_0$, where k_0 is the free-space wavenumber, $k_0 = \omega\sqrt{\mu_0\epsilon_0}$. Setting also $n = k/k_0 = \sqrt{\mu\epsilon}/\sqrt{\mu_0\epsilon_0}$, we have:

$$k_\pm = n_\pm k_0\,, \quad n_\pm = n(1 \pm a) \tag{4.3.7}$$

For the $\{E_-, H_-\}$ circular components, we define the left-polarized (forward-moving) and right-polarized (backward-moving) fields by:

$$\begin{aligned} E_{L-} &= \frac{1}{2}[E_- + j\eta H_-] \\ E_{R-} &= \frac{1}{2}[E_- - j\eta H_-] \end{aligned} \quad\Leftrightarrow\quad \begin{aligned} E_- &= E_{L-} + E_{R-} \\ H_- &= \frac{1}{j\eta}[E_{L-} - E_{R-}] \end{aligned} \tag{4.3.8}$$

Then, $\{E_{L-}, E_{R-}\}$ will satisfy:

$$\frac{\partial}{\partial z}\begin{bmatrix} E_{L-} \\ E_{R-} \end{bmatrix} = \begin{bmatrix} -jk_- & 0 \\ 0 & jk_+ \end{bmatrix}\begin{bmatrix} E_{L-} \\ E_{R-} \end{bmatrix} \quad\Rightarrow\quad \begin{aligned} E_{L-}(z) &= A_- e^{-jk_- z} \\ E_{R-}(z) &= B_- e^{jk_+ z} \end{aligned} \tag{4.3.9}$$

In summary, we obtain the complete circular-basis fields $E_\pm(z)$:

$$\boxed{\begin{aligned} E_+(z) &= E_{R+}(z) + E_{L+}(z) = A_+ e^{-jk_+ z} + B_+ e^{jk_- z} \\ E_-(z) &= E_{L-}(z) + E_{R-}(z) = A_- e^{-jk_- z} + B_- e^{jk_+ z} \end{aligned}} \tag{4.3.10}$$

Thus, the $E_+(z)$ circular component propagates forward with wavenumber k_+ and backward with k_-, and the reverse is true of the $E_-(z)$ component. The forward-moving component of E_+ and the backward-moving component of E_-, that is, E_{R+} and E_{R-}, are both right-polarized and both propagate with the same wavenumber k_+. Similarly, the left-polarized waves E_{L+} and E_{L-} both propagate with k_-.

Thus, a wave of given circular polarization (left or right) propagates with the same wavenumber regardless of its direction of propagation. This is a characteristic difference of chiral versus gyrotropic media in external magnetic fields.

Consider, next, the effect of natural rotation. We start with a linearly polarized field at $z = 0$ and decompose it into its circular components:

$$\boldsymbol{E}(0) = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y = \hat{\mathbf{e}}_+A_+ + \hat{\mathbf{e}}_-A_- \,, \quad \text{with} \quad A_\pm = \frac{1}{2}(A_x \pm jA_y)$$

where A_x, A_y must be real for linear polarization. Propagating the circular components forward by a distance l according to Eq. (4.3.10), we find:

$$\begin{aligned} \boldsymbol{E}(l) &= \hat{\mathbf{e}}_+A_+\,e^{-jk_+l} + \hat{\mathbf{e}}_-A_-\,e^{-jk_-l} \\ &= \big[\hat{\mathbf{e}}_+A_+e^{-j(k_+-k_-)l/2} + \hat{\mathbf{e}}_-A_-e^{j(k_+-k_-)l/2}\big]e^{-j(k_++k_-)l/2} \\ &= \big[\hat{\mathbf{e}}_+A_+e^{-j\phi} + \hat{\mathbf{e}}_-A_-e^{j\phi}\big]e^{-j(k_++k_-)l/2} \end{aligned} \tag{4.3.11}$$

where we defined the angle of rotation:

$$\boxed{\phi = \frac{1}{2}(k_+ - k_-)l = akl} \qquad \text{(natural rotation)} \tag{4.3.12}$$

Going back to the linear basis, we find:

$$\begin{aligned} \hat{\mathbf{e}}_+A_+e^{-j\phi} + \hat{\mathbf{e}}_-A_-e^{j\phi} &= (\hat{\mathbf{x}} - j\hat{\mathbf{y}})\frac{1}{2}(A_x + jA_y)\,e^{-j\phi} + (\hat{\mathbf{x}} + j\hat{\mathbf{y}})\frac{1}{2}(A_x - jA_y)\,e^{j\phi} \\ &= \big[\hat{\mathbf{x}}\cos\phi - \hat{\mathbf{y}}\sin\phi\big]A_x + \big[\hat{\mathbf{y}}\cos\phi + \hat{\mathbf{x}}\sin\phi\big]A_y \\ &= \hat{\mathbf{x}}'A_x + \hat{\mathbf{y}}'A_y \end{aligned}$$

Therefore, at $z = 0$ and $z = l$, we have:

$$\begin{aligned} \boldsymbol{E}(0) &= \big[\hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y\big] \\ \boldsymbol{E}(l) &= \big[\hat{\mathbf{x}}'A_x + \hat{\mathbf{y}}'A_y\big]e^{-j(k_++k_-)l/2} \end{aligned} \tag{4.3.13}$$

The new unit vectors $\hat{\mathbf{x}}' = \hat{\mathbf{x}}\cos\phi - \hat{\mathbf{y}}\sin\phi$ and $\hat{\mathbf{y}}' = \hat{\mathbf{y}}\cos\phi + \hat{\mathbf{x}}\sin\phi$ are recognized as the unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ rotated clockwise (if $\phi > 0$) by the angle ϕ, as shown in Fig. 4.2.1 (for the case $A_x \neq 0$, $A_y = 0$.) Thus, the wave remains linearly polarized, but its polarization plane rotates as it propagates.

If the propagation is in the negative z-direction, then as follows from Eq. (4.3.10), the roles of k_+ and k_- are interchanged so that the rotation angle becomes $\phi = (k_- - k_+)l/2$, which is the negative of that of Eq. (4.3.12).

If a linearly polarized wave travels forward by a distance l, gets reflected, and travels back to the starting point, the total angle of rotation will be zero. By contrast, in the Faraday rotation case, the angle keeps increasing so that it doubles after a round trip (see Problem 4.10.)

4.4 Gyrotropic Media

Gyrotropic[†] media are isotropic media in the presence of constant external magnetic fields. A gyroelectric medium (at frequency ω) has constitutive relationships:

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_1 & j\epsilon_2 & 0 \\ -j\epsilon_2 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}, \quad \boldsymbol{B} = \mu\boldsymbol{H} \tag{4.4.1}$$

For a lossless medium, the positivity of the energy density function requires that the permittivity matrix be hermitian and positive-definite, which implies that $\epsilon_1, \epsilon_2, \epsilon_3$ are real, and moreover, $\epsilon_1 > 0$, $|\epsilon_2| \leq \epsilon_1$, and $\epsilon_3 > 0$. The quantity ϵ_2 is proportional to the external magnetic field and reverses sign with the direction of that field.

A gyromagnetic medium, such as a ferrite in the presence of a magnetic field, has similar constitutive relationships, but with the roles of $\boldsymbol{D}$ and $\boldsymbol{H}$ interchanged:

$$\begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} = \begin{bmatrix} \mu_1 & j\mu_2 & 0 \\ -j\mu_2 & \mu_1 & 0 \\ 0 & 0 & \mu_3 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}, \quad \boldsymbol{D} = \epsilon\boldsymbol{E} \tag{4.4.2}$$

where again $\mu_1 > 0$, $|\mu_2| \leq \mu_1$, and $\mu_3 > 0$ for a lossless medium.

In the circular basis of Eq. (4.1.1), the above gyrotropic constitutive relationships take the simplified forms:

$$\begin{aligned} D_\pm &= (\epsilon_1 \pm \epsilon_2)E_\pm \,, \quad & B_\pm &= \mu H_\pm \,, \quad & &\text{(gyroelectric)} \\ B_\pm &= (\mu_1 \pm \mu_2)H_\pm \,, \quad & D_\pm &= \epsilon E_\pm \,, \quad & &\text{(gyromagnetic)} \end{aligned} \tag{4.4.3}$$

where we ignored the z-components, which are zero for a uniform plane wave propagating in the z-direction. For example,

$$D_x \pm jD_y = (\epsilon_1 E_x + j\epsilon_2 E_y) \pm j(\epsilon_1 E_y - j\epsilon_2 E_x) = (\epsilon_1 \pm \epsilon_2)\,(E_x \pm jE_y)$$

Next, we solve Eqs. (4.1.5) for the forward and backward circular-basis waves. Considering the gyroelectric case first, we define the following quantities:

$$\boxed{\epsilon_\pm = \epsilon_1 \pm \epsilon_2 \,, \quad k_\pm = \omega\sqrt{\mu\epsilon_\pm} \,, \quad \eta_\pm = \sqrt{\frac{\mu}{\epsilon_\pm}}} \tag{4.4.4}$$

Using these definitions and the constitutive relations $D_\pm = \epsilon_\pm E_\pm$, Eqs. (4.1.5) may be rearranged into the following matrix forms:

$$\frac{\partial}{\partial z}\begin{bmatrix} E_\pm \\ \eta_\pm H_\pm \end{bmatrix} = \begin{bmatrix} 0 & \mp k_\pm \\ \pm k_\pm & 0 \end{bmatrix} \begin{bmatrix} E_\pm \\ \eta_\pm H_\pm \end{bmatrix} \tag{4.4.5}$$

These may be decoupled by defining forward- and backward-moving fields as in Eqs. (4.3.4) and (4.3.8), but using the corresponding circular impedances $\eta_\pm$:

$$\begin{aligned} E_{R+} &= \frac{1}{2}\big[E_+ - j\eta_+H_+\big] \qquad & E_{L-} &= \frac{1}{2}\big[E_- + j\eta_-H_-\big] \\ E_{L+} &= \frac{1}{2}\big[E_+ + j\eta_+H_+\big] \qquad & E_{R-} &= \frac{1}{2}\big[E_- - j\eta_-H_-\big] \end{aligned} \tag{4.4.6}$$

[†] The term “gyrotropic” is sometimes also used to mean “optically active.”

These satisfy the decoupled equations:

$$\begin{aligned}
&\frac{\partial}{\partial z}\begin{bmatrix} E_{R+} \\ E_{L+} \end{bmatrix} = \begin{bmatrix} -jk_+ & 0 \\ 0 & jk_+ \end{bmatrix}\begin{bmatrix} E_{R+} \\ E_{L+} \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} E_{R+}(z) &= A_+\, e^{-jk_+ z} \\ E_{L+}(z) &= B_+\, e^{jk_+ z} \end{aligned} \\
&\frac{\partial}{\partial z}\begin{bmatrix} E_{L-} \\ E_{R-} \end{bmatrix} = \begin{bmatrix} -jk_- & 0 \\ 0 & jk_- \end{bmatrix}\begin{bmatrix} E_{L-} \\ E_{R-} \end{bmatrix} \quad \Rightarrow \quad \begin{aligned} E_{L-}(z) &= A_-\, e^{-jk_- z} \\ E_{R-}(z) &= B_-\, e^{jk_- z} \end{aligned}
\end{aligned} \tag{4.4.7}$$

Thus, the complete circular-basis fields $E_\pm(z)$ are:

$$\boxed{\begin{aligned} E_+(z) &= E_{R+}(z) + E_{L+}(z) = A_+\, e^{-jk_+ z} + B_+\, e^{jk_+ z} \\ E_-(z) &= E_{L-}(z) + E_{R-}(z) = A_-\, e^{-jk_- z} + B_-\, e^{jk_- z} \end{aligned}} \tag{4.4.8}$$

Now, the $E_+(z)$ circular component propagates forward and backward with the *same* wavenumber k_+, while $E_-(z)$ propagates with k_-. Eq. (4.3.13) and the steps leading to it remain valid here. The rotation of the polarization plane is referred to as the *Faraday rotation*. If the propagation is in the negative z-direction, then the roles of k_+ and k_- remain unchanged so that the rotation angle is still the same as that of Eq. (4.3.12).

If a linearly polarized wave travels forward by a distance l, gets reflected, and travels back to the starting point, the total angle of rotation will be double that of the single trip, that is, $2\phi = (k_+ - k_-)l$.

Problems 1.10 and 4.12 discuss simple models of gyroelectric behavior for conductors and plasmas in the presence of an external magnetic field. Problem 4.14 develops the Appleton-Hartree formulas for plane waves propagating in plasmas, such as the ionosphere [696–700].

The gyromagnetic case is essentially identical to the gyroelectric one. Eqs. (4.4.5) to (4.4.8) remain the same, but with circular wavenumbers and impedances defined by:

$$\boxed{\mu_\pm = \mu_1 \pm \mu_2\,, \quad k_\pm = \omega\sqrt{\epsilon\mu_\pm}\,, \quad \eta_\pm = \sqrt{\frac{\mu_\pm}{\epsilon}}} \tag{4.4.9}$$

Problem 4.13 discusses a model for magnetic resonance exhibiting gyromagnetic behavior. Magnetic resonance has many applications—from NMR imaging to ferrite microwave devices [701–712]. Historical overviews may be found in [710,712].

4.5 Linear and Circular Dichroism

Dichroic polarizers, such as polaroids, are linearly birefringent materials that have widely different attenuation coefficients along the two polarization directions. For a lossy material, the field solutions given in Eq. (4.2.4) are modified as follows:

$$\begin{aligned} E_x(z) &= Ae^{-jk_1 z} = Ae^{-\alpha_1 z}e^{-j\beta_1 z}\,, \qquad & k_1 &= \omega\sqrt{\mu\epsilon_1} = \beta_1 - j\alpha_1 \\ E_y(z) &= Be^{-jk_2 z} = Be^{-\alpha_2 z}e^{-j\beta_1 2}\,, \qquad & k_2 &= \omega\sqrt{\mu\epsilon_2} = \beta_2 - j\alpha_2 \end{aligned} \tag{4.5.1}$$

where α_1, α_2 are the attenuation coefficients. Passing through a length l of such a material, the initial and output polarizations will be as follows:

$$\begin{aligned} \mathbf{E}(0) &= \hat{\mathbf{x}}\,A + \hat{\mathbf{y}}\,B \\ \mathbf{E}(l) &= \hat{\mathbf{x}}\,Ae^{-jk_1 l} + \hat{\mathbf{y}}\,Be^{-jk_2 l} = \big(\hat{\mathbf{x}}\,Ae^{-\alpha_1 l} + \hat{\mathbf{y}}\,Be^{-\alpha_2 l}e^{j\phi}\big)e^{-j\beta_1 l} \end{aligned} \tag{4.5.2}$$

In addition to the phase change $\phi = (\beta_1 - \beta_2)l$, the field amplitudes have attenuated by the unequal factors $a_1 = e^{-\alpha_1 l}$ and $a_2 = e^{-\alpha_2 l}$. The resulting polarization will be elliptic with unequal semi-axes. If $\alpha_2 \gg \alpha_1$, then $a_2 \ll a_1$ and the y-component can be ignored in favor of the x-component.

This is the basic principle by which a polaroid material lets through only a preferred linear polarization. An ideal linear polarizer would have $a_1 = 1$ and $a_2 = 0$, corresponding to $\alpha_1 = 0$ and $\alpha_2 = \infty$. Typical values of the attenuations for commercially available polaroids are of the order of $a_1 = 0.9$ and $a_2 = 10^{-2}$, or 0.9 dB and 40 dB, respectively.

Chiral media may exhibit *circular dichroism* [681,694], in which the circular wavenumbers become complex, $k_\pm = \beta_\pm - j\alpha_\pm$. Eq. (4.3.11) reads now:

$$\begin{aligned} \mathbf{E}(l) &= \hat{\mathbf{e}}_+ A_+\, e^{-jk_+ l} + \hat{\mathbf{e}}_- A_-\, e^{-jk_- l} \\ &= \big[\hat{\mathbf{e}}_+ A_+ e^{-j(k_+ - k_-)l/2} + \hat{\mathbf{e}}_- A_- e^{j(k_+ - k_-)l/2}\big]e^{-j(k_+ + k_-)l/2} \\ &= \big[\hat{\mathbf{e}}_+ A_+ e^{-\psi - j\phi} + \hat{\mathbf{e}}_- A_- e^{\psi + j\phi}\big]e^{-j(k_+ + k_-)l/2} \end{aligned} \tag{4.5.3}$$

where we defined the complex rotation angle:

$$\phi - j\psi = \frac{1}{2}(k_+ - k_-)l = \frac{1}{2}(\beta_+ - \beta_-)l - j\,\frac{1}{2}(\alpha_+ - \alpha_-)l \tag{4.5.4}$$

Going back to the linear basis as in Eq. (4.3.13), we obtain:

$$\begin{aligned} \mathbf{E}(0) &= \big[\hat{\mathbf{x}}\,A_x + \hat{\mathbf{y}}\,A_y\big] \\ \mathbf{E}(l) &= \big[\hat{\mathbf{x}}'A_x' + \hat{\mathbf{y}}'A_y'\big]e^{-j(k_+ + k_-)l/2} \end{aligned} \tag{4.5.5}$$

where $\{\hat{\mathbf{x}}', \hat{\mathbf{y}}'\}$ are the same rotated (by ϕ) unit vectors of Eq. (4.3.13), and

$$\begin{aligned} A_x' &= A_x \cosh\psi - jA_y \sinh\psi \\ A_y' &= A_y \cosh\psi + jA_x \sinh\psi \end{aligned} \tag{4.5.6}$$

Because the amplitudes A_x', A_y' are now complex-valued, the resulting polarization will be elliptical.

4.6 Oblique Propagation in Birefringent Media

Here, we discuss TE and TM waves propagating in oblique directions in linearly birefringent media. We will use these results in Chap. 8 to discuss reflection and refraction in such media, and to characterize the properties of birefringent multilayer structures.

Applications include the recently manufactured (by 3M, Inc.) multilayer birefringent polymer mirrors that have remarkable and unusual optical properties, collectively referred to as *giant birefringent optics* (GBO) [658].

Oblique propagation in chiral and gyrotropic media is discussed in the problems. Further discussions of wave propagation in anisotropic media may be found in [30–32].

We recall from Sec. 2.9 that a uniform plane wave propagating in a lossless isotropic dielectric in the direction of a wave vector $\boldsymbol{k}$ is given by:

$$\boldsymbol{E}(\mathbf{r}) = \boldsymbol{E}e^{-j\boldsymbol{k}\cdot\mathbf{r}}, \quad \boldsymbol{H}(\mathbf{r}) = \boldsymbol{H}e^{-j\boldsymbol{k}\cdot\mathbf{r}}, \quad \text{with} \quad \hat{\mathbf{k}}\cdot\boldsymbol{E} = 0, \quad \boldsymbol{H} = \frac{n}{\eta_0}\hat{\mathbf{k}}\times\boldsymbol{E} \tag{4.6.1}$$

where n is the refractive index of the medium $n = \sqrt{\epsilon/\epsilon_0}$, η_0 the free-space impedance, and $\hat{\mathbf{k}}$ the unit-vector in the direction of $\boldsymbol{k}$, so that,

$$\boldsymbol{k} = k\,\hat{\mathbf{k}}, \quad k = |\boldsymbol{k}| = \omega\sqrt{\mu_0\epsilon} = nk_0, \quad k_0 = \frac{\omega}{c_0} = \omega\sqrt{\mu_0\epsilon_0} \tag{4.6.2}$$

and k_0 is the free-space wavenumber. Thus, $\boldsymbol{E}, \boldsymbol{H}, \hat{\mathbf{k}}$ form a right-handed system.

In particular, following the notation of Fig. 2.9.1, if $\boldsymbol{k}$ is chosen to lie in the xz plane at an angle θ from the z-axis, that is, $\hat{\mathbf{k}} = \hat{\mathbf{x}}\sin\theta + \hat{\mathbf{z}}\cos\theta$, then there will be two independent polarization solutions: TM, parallel, or p-polarization, and TE, perpendicular, or s-polarization, with fields given by

$$\begin{aligned} &\text{(TM, p-polarization):} \qquad \boldsymbol{E} = E_0(\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta), \quad \boldsymbol{H} = \frac{n}{\eta_0}E_0\,\hat{\mathbf{y}} \\ &\text{(TE, s-polarization):} \qquad \boldsymbol{E} = E_0\,\hat{\mathbf{y}}, \quad \boldsymbol{H} = \frac{n}{\eta_0}E_0(-\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{z}}\sin\theta) \end{aligned} \tag{4.6.3}$$

where, in both the TE and TM cases, the propagation phase factor $e^{-j\boldsymbol{k}\cdot\mathbf{r}}$ is:

$$e^{-j\boldsymbol{k}\cdot\mathbf{r}} = e^{-j(k_z z + k_x x)} = e^{-jk_0 n(z\cos\theta + x\sin\theta)} \tag{4.6.4}$$

The designation as parallel or perpendicular is completely arbitrary here and is taken with respect to the xz plane. In the reflection and refraction problems discussed in Chap. 7, the dielectric interface is taken to be the xy plane and the xz plane becomes the plane of incidence.

In a birefringent medium, the propagation of a uniform plane wave with arbitrary wave vector $\boldsymbol{k}$ is much more difficult to describe. For example, the direction of the Poynting vector is not towards $\boldsymbol{k}$, the electric field $\boldsymbol{E}$ is not orthogonal to $\boldsymbol{k}$, the simple dispersion relationship $k = n\omega/c_0$ is not valid, and so on.

In the previous section, we considered the special case of propagation along an ordinary optic axis in a birefringent medium. Here, we discuss the somewhat more general case in which the xyz coordinate axes coincide with the principal dielectric axes (so that the permittivity tensor is diagonal,) and we take the wave vector $\boldsymbol{k}$ to lie in the xz plane at an angle θ from the z-axis. The geometry is depicted in Fig. 4.6.1.

Although this case is still not the most general one with a completely arbitrary direction for $\boldsymbol{k}$, it does contain most of the essential features of propagation in birefringent media. The 3M multilayer films mentioned above have similar orientations for their optic axes [658].

The constitutive relations are assumed to be $\boldsymbol{B} = \mu_0\boldsymbol{H}$ and a diagonal permittivity tensor for $\boldsymbol{D}$. Let $\epsilon_1, \epsilon_2, \epsilon_3$ be the permittivity values along the three principal axes and define the corresponding refractive indices $n_i = \sqrt{\epsilon_i/\epsilon_0}$, $i = 1, 2, 3$. Then, the $\boldsymbol{D}$-$\boldsymbol{E}$ relationship becomes:

$$\begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \epsilon_0 \begin{bmatrix} n_1^2 & 0 & 0 \\ 0 & n_2^2 & 0 \\ 0 & 0 & n_3^2 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{4.6.5}$$

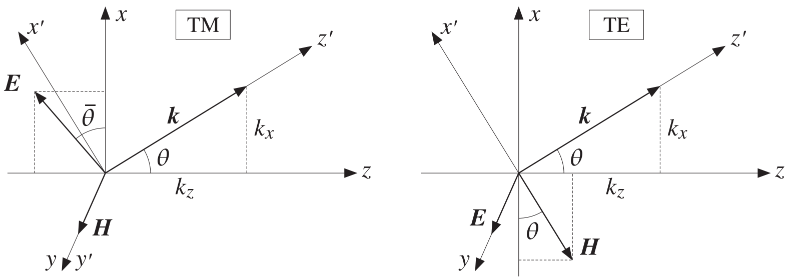


Fig. 4.6.1 Uniform plane waves in a birefringent medium.

For a biaxial medium, the three n_i are all different. For a uniaxial medium, we take the xy-axes to be *ordinary*, with $n_1 = n_2 = n_o$, and the z-axis to be *extraordinary*, with $n_3 = n_e$.[†] The wave vector $\boldsymbol{k}$ can be resolved along the z and x directions as follows:

$$\boldsymbol{k} = k\,\hat{\mathbf{k}} = k(\hat{\mathbf{x}}\sin\theta + \hat{\mathbf{z}}\cos\theta) = \hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_z \tag{4.6.6}$$

The ω-k relationship is determined from the solution of Maxwell's equations. By analogy with the isotropic case that has $k = nk_0 = n\omega/c_0$, we may define an *effective* refractive index N such that:

$$\boxed{k = Nk_0 = N\frac{\omega}{c_0}} \qquad \text{(effective refractive index)} \tag{4.6.7}$$

We will see in Eq. (4.6.22) by solving Maxwell's equations that N depends on the chosen polarization (according to Fig. 4.6.1) and on the wave vector direction θ:

$$N = \begin{cases} \dfrac{n_1 n_3}{\sqrt{n_1^2\sin^2\theta + n_3^2\cos^2\theta}}, & \text{(TM, p-polarization)} \\ n_2, & \text{(TE, s-polarization)} \end{cases} \tag{4.6.8}$$

For the TM case, we may rewrite the N-θ relationship in the form:

$$\boxed{\frac{1}{N^2} = \frac{\cos^2\theta}{n_1^2} + \frac{\sin^2\theta}{n_3^2}} \qquad \text{(effective TM index)} \tag{4.6.9}$$

Multiplying by k^2 and using $k_0 = k/N$, and $k_x = k\sin\theta$, $k_z = k\cos\theta$, we obtain the ω-k relationship for the TM case:

$$\frac{\omega^2}{c_0^2} = \frac{k_z^2}{n_1^2} + \frac{k_x^2}{n_3^2} \qquad \text{(TM, p-polarization)} \tag{4.6.10}$$

Similarly, we have for the TE case:

$$\frac{\omega^2}{c_0^2} = \frac{k^2}{n_2^2} \qquad \text{(TE, s-polarization)} \tag{4.6.11}$$

[†] In Sec. 4.2, the extraordinary axis was the x-axis.

Thus, the TE mode propagates as if the medium were isotropic with index $n = n_2$, whereas the TM mode propagates in a more complicated fashion. If the wave vector $\boldsymbol{k}$ is along the ordinary x-axis ($\theta = 90^{\rm o}$), then $k = k_x = n_3\omega/c_0$ (this was the result of the previous section), and if $\boldsymbol{k}$ is along the extraordinary z-axis ($\theta = 0^{\rm o}$), then we have $k = k_z = n_1\omega/c_0$.

For TM modes, the group velocity is not along $\boldsymbol{k}$. In general, the group velocity depends on the ω-k relationship and is computed as $\boldsymbol{v} = \partial\omega/\partial\boldsymbol{k}$. From Eq. (4.6.10), we find the x- and z-components:

$$
\begin{aligned}
v_x &= \frac{\partial\omega}{\partial k_x} = \frac{k_x c_0^2}{\omega n_3^2} = c_0\,\frac{N}{n_3^2}\,\sin\theta \\
v_z &= \frac{\partial\omega}{\partial k_z} = \frac{k_z c_0^2}{\omega n_1^2} = c_0\,\frac{N}{n_1^2}\,\cos\theta
\end{aligned}
\tag{4.6.12}
$$

The velocity vector $\boldsymbol{v}$ is not parallel to $\boldsymbol{k}$. The angle $\bar\theta$ that $\boldsymbol{v}$ forms with the z-axis is given by $\tan\bar\theta = v_x/v_z$. It follows from (4.6.12) that:

$$
\boxed{\tan\bar\theta = \frac{n_1^2}{n_3^2}\,\tan\theta} \qquad \text{(group velocity direction)} \tag{4.6.13}
$$

Clearly, $\bar\theta \neq \theta$ if $n_1 \neq n_3$. The relative directions of $\boldsymbol{k}$ and $\boldsymbol{v}$ are shown in Fig. 4.6.2. The group velocity is also equal to the energy transport velocity defined in terms of the Poynting vector $\boldsymbol{\mathcal{P}}$ and energy density w as $\boldsymbol{v} = \boldsymbol{\mathcal{P}}/w$. Thus, $\boldsymbol{v}$ and $\boldsymbol{\mathcal{P}}$ have the same direction. Moreover, with the electric field being orthogonal to the Poynting vector, the angle $\bar\theta$ is also equal to the angle the $\boldsymbol{E}$-field forms with the x-axis.

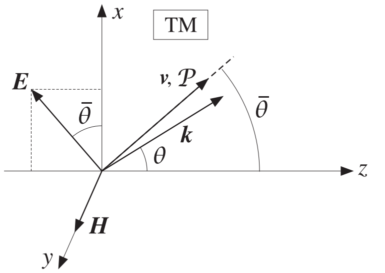


Fig. 4.6.2 Directions of group velocity, Poynting vector, wave vector, and electric field.

Next, we derive Eqs. (4.6.8) for N and solve for the field components in the TM and TE cases. We look for propagating solutions of Maxwell's equations of the type $\boldsymbol{E}(\mathbf{r}) = \boldsymbol{E}e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}$ and $\boldsymbol{H}(\mathbf{r}) = \boldsymbol{H}e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}$. Replacing the gradient operator by $\nabla \to -j\boldsymbol{k}$ and canceling some factors of j, Maxwell's equations take the form:

$$
\begin{aligned}
\nabla\times\boldsymbol{E} &= -j\omega\mu_0\boldsymbol{H} & \qquad & \boldsymbol{k}\times\boldsymbol{E} = \omega\mu_0\boldsymbol{H} \\
\nabla\times\boldsymbol{H} &= j\omega\boldsymbol{D} & \Rightarrow \qquad & \boldsymbol{k}\times\boldsymbol{H} = -\omega\boldsymbol{D} \\
\nabla\cdot\boldsymbol{D} &= 0 & & \boldsymbol{k}\cdot\boldsymbol{D} = 0 \\
\nabla\cdot\boldsymbol{H} &= 0 & & \boldsymbol{k}\cdot\boldsymbol{H} = 0
\end{aligned}
\tag{4.6.14}
$$

The last two equations are implied by the first two, as can be seen by dotting both sides of the first two with $\boldsymbol{k}$. Replacing $\boldsymbol{k} = k\,\hat{\mathbf{k}} = Nk_0\hat{\mathbf{k}}$, where N is still to be determined, we may solve Faraday's law for $\boldsymbol{H}$ in terms of $\boldsymbol{E}$:

$$
N\,\frac{\omega}{c_0}\,\hat{\mathbf{k}}\times\boldsymbol{E} = \omega\mu_0\boldsymbol{H} \quad\Rightarrow\quad \boxed{\boldsymbol{H} = \frac{N}{\eta_0}\,\hat{\mathbf{k}}\times\boldsymbol{E}} \tag{4.6.15}
$$

where we used $\eta_0 = c_0\mu_0$. Then, Ampère's law gives:

$$
\boldsymbol{D} = -\frac{1}{\omega}\,\boldsymbol{k}\times\boldsymbol{H} = -\frac{1}{\omega}\,N\frac{\omega}{c_0}\,\hat{\mathbf{k}}\times\boldsymbol{H} = \frac{N^2}{\eta_0 c_0}\,\hat{\mathbf{k}}\times(\boldsymbol{E}\times\hat{\mathbf{k}}) \quad\Rightarrow\quad \boxed{\hat{\mathbf{k}}\times(\boldsymbol{E}\times\hat{\mathbf{k}}) = \frac{1}{\epsilon_0 N^2}\,\boldsymbol{D}}
$$

where we used $c_0\eta_0 = 1/\epsilon_0$. The quantity $\hat{\mathbf{k}}\times(\boldsymbol{E}\times\hat{\mathbf{k}})$ is recognized as the component of $\boldsymbol{E}$ that is transverse to the propagation unit vector $\hat{\mathbf{k}}$. Using the BAC-CAB vector identity, we have $\hat{\mathbf{k}}\times(\boldsymbol{E}\times\hat{\mathbf{k}}) = \boldsymbol{E} - \hat{\mathbf{k}}(\hat{\mathbf{k}}\cdot\boldsymbol{E})$. Rearranging terms, we obtain:

$$
\boxed{\boldsymbol{E} - \frac{1}{\epsilon_0 N^2}\,\boldsymbol{D} = \hat{\mathbf{k}}(\hat{\mathbf{k}}\cdot\boldsymbol{E})} \tag{4.6.16}
$$

Because $\boldsymbol{D}$ is linear in $\boldsymbol{E}$, this is a homogeneous linear equation. Therefore, in order to have a nonzero solution, its determinant must be zero. This provides a condition from which N can be determined.

To obtain both the TE and TM solutions, we assume initially that $\boldsymbol{E}$ has all its three components and rewrite Eq. (4.6.16) component-wise. Using Eq. (4.6.5) and noting that $\hat{\mathbf{k}}\cdot\boldsymbol{E} = E_x\sin\theta + E_z\cos\theta$, we obtain the homogeneous linear system:

$$
\begin{aligned}
\Big(1 - \frac{n_1^2}{N^2}\Big)E_x &= (E_x\sin\theta + E_z\cos\theta)\sin\theta \\
\Big(1 - \frac{n_2^2}{N^2}\Big)E_y &= 0 \\
\Big(1 - \frac{n_3^2}{N^2}\Big)E_z &= (E_x\sin\theta + E_z\cos\theta)\cos\theta
\end{aligned}
\tag{4.6.17}
$$

The TE case has $E_y \neq 0$ and $E_x = E_z = 0$, whereas the TM case has $E_x \neq 0$, $E_z \neq 0$, and $E_y = 0$. Thus, the two cases decouple.

In the TE case, the second of Eqs. (4.6.17) immediately implies that $N = n_2$. Setting $\boldsymbol{E} = E_0\hat{\mathbf{y}}$ and using $\hat{\mathbf{k}}\times\hat{\mathbf{y}} = -\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{z}}\sin\theta$, we obtain the TE solution:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= E_0\,\hat{\mathbf{y}}\,e^{-j\boldsymbol{k}\cdot\mathbf{r}} \\ \boldsymbol{H}(\mathbf{r}) &= \frac{n_2}{\eta_0}\,E_0\,(-\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{z}}\sin\theta)\,e^{-j\boldsymbol{k}\cdot\mathbf{r}} \end{aligned}} \quad \text{(TE)} \qquad (4.6.18)$$

where the TE propagation phase factor is:

$$\boxed{e^{-j\boldsymbol{k}\cdot\mathbf{r}} = e^{-jk_0 n_2(z\cos\theta + x\sin\theta)}} \quad \text{(TE propagation factor)} \qquad (4.6.19)$$

The TM case requires a little more work. The linear system (4.6.17) becomes now:

$$\begin{aligned} \left(1 - \frac{n_1^2}{N^2}\right)E_x &= (E_x\sin\theta + E_z\cos\theta)\sin\theta \\ \left(1 - \frac{n_3^2}{N^2}\right)E_z &= (E_x\sin\theta + E_z\cos\theta)\cos\theta \end{aligned} \qquad (4.6.20)$$

Using the identity $\sin^2\theta + \cos^2\theta = 1$, we may rewrite Eq. (4.6.20) in the matrix form:

$$\begin{bmatrix} \cos^2\theta - \dfrac{n_1^2}{N^2} & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin^2\theta - \dfrac{n_3^2}{N^2} \end{bmatrix} \begin{bmatrix} E_x \\ E_z \end{bmatrix} = 0 \qquad (4.6.21)$$

Setting the determinant of the coefficient matrix to zero, we obtain the desired condition on N in order that a non-zero solution E_x, E_z exist:

$$\left(\cos^2\theta - \frac{n_1^2}{N^2}\right)\left(\sin^2\theta - \frac{n_3^2}{N^2}\right) - \sin^2\theta\cos^2\theta = 0 \qquad (4.6.22)$$

This can be solved for N^2 to give Eq. (4.6.9). From it, we may also derive the following relationship, which will prove useful in applying Snel's law in birefringent media:

$$\boxed{N\cos\theta = \frac{n_1}{n_3}\sqrt{n_3^2 - N^2\sin^2\theta} = n_1\sqrt{1 - \frac{N^2\sin^2\theta}{n_3^2}}} \qquad (4.6.23)$$

With the help of the relationships given in Problem 4.16, the solution of the homogeneous system (4.6.20) is found to be, up to a proportionality constant:

$$E_x = A\,\frac{n_3}{n_1}\cos\theta\,, \quad E_z = -A\,\frac{n_1}{n_3}\sin\theta \qquad (4.6.24)$$

The constant A can be expressed in terms of the total magnitude of the field $E_0 = |\boldsymbol{E}| = \sqrt{|E_x|^2 + |E_z|^2}$. Using the relationship (4.7.11), we find (assuming $A > 0$):

$$A = E_0\,\frac{N}{\sqrt{n_1^2 + n_3^2 - N^2}} \qquad (4.6.25)$$

The magnetic field $\boldsymbol{H}$ can also be expressed in terms of the constant A. We have:

$$\begin{aligned} \boldsymbol{H} &= \frac{N}{\eta_0}\,\hat{\mathbf{k}}\times\boldsymbol{E} = \frac{N}{\eta_0}\,(\hat{\mathbf{x}}\sin\theta + \hat{\mathbf{z}}\cos\theta)\times(\hat{\mathbf{x}}E_x + \hat{\mathbf{z}}E_z) \\ &= \frac{N}{\eta_0}\,\hat{\mathbf{y}}(E_x\cos\theta - E_z\sin\theta) = \frac{N}{\eta_0}\,\hat{\mathbf{y}}\,A\left(\frac{n_3}{n_1}\cos^2\theta + \frac{n_1}{n_3}\sin^2\theta\right) \\ &= \frac{N}{\eta_0}\,\hat{\mathbf{y}}\,A\,\frac{n_1 n_3}{N^2} = \frac{A}{\eta_0}\,\hat{\mathbf{y}}\,\frac{n_1 n_3}{N} \end{aligned} \qquad (4.6.26)$$

where we used Eq. (4.7.10). In summary, the complete TM solution is:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= E_0\,\frac{N}{\sqrt{n_1^2 + n_3^2 - N^2}}\left(\hat{\mathbf{x}}\,\frac{n_3}{n_1}\cos\theta - \hat{\mathbf{z}}\,\frac{n_1}{n_3}\sin\theta\right)e^{-j\boldsymbol{k}\cdot\mathbf{r}} \\ \boldsymbol{H}(\mathbf{r}) &= \frac{E_0}{\eta_0}\,\frac{n_1 n_3}{\sqrt{n_1^2 + n_3^2 - N^2}}\,\hat{\mathbf{y}}\,e^{-j\boldsymbol{k}\cdot\mathbf{r}} \end{aligned}} \quad \text{(TM)} \qquad (4.6.27)$$

where the TM propagation phase factor is:

$$\boxed{e^{-j\boldsymbol{k}\cdot\mathbf{r}} = e^{-jk_0 N(z\cos\theta + x\sin\theta)}} \quad \text{(TM propagation factor)} \qquad (4.6.28)$$

The solution has been put in a form that exhibits the proper limits at $\theta = 0^{\mathrm{o}}$ and 90^{o}. It agrees with Eq. (4.6.3) in the isotropic case. The angle that $\boldsymbol{E}$ forms with the x-axis in Fig. 4.6.2 is given by $\tan\bar{\theta} = -E_z/E_x$ and agrees with Eq. (4.6.13).

Next, we derive expressions for the Poynting vector and energy densities. It turns out—as is common in propagation and waveguide problems—that the magnetic energy density is equal to the electric one. Using Eq. (4.6.27), we find:

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\,\mathrm{Re}(\boldsymbol{E}\times\boldsymbol{H}^*) = \frac{E_0^2}{2\eta_0}\,\frac{n_1 n_3 N}{n_1^2 + n_3^2 - N^2}\left(\hat{\mathbf{x}}\,\frac{n_1}{n_3}\sin\theta + \hat{\mathbf{z}}\,\frac{n_3}{n_1}\cos\theta\right) \qquad (4.6.29)$$

and for the electric, magnetic, and total energy densities:

$$\begin{aligned} w_e &= \frac{1}{2}\,\mathrm{Re}(\boldsymbol{D}\cdot\boldsymbol{E}^*) = \frac{1}{4}\epsilon_0\big(n_1^2|E_x|^2 + n_3^2|E_z|^2\big) = \frac{1}{4}\epsilon_0 E_0^2\,\frac{n_1^2 n_3^2}{n_1^2 + n_3^2 - N^2} \\ w_m &= \frac{1}{2}\,\mathrm{Re}(\boldsymbol{B}\cdot\boldsymbol{H}^*) = \frac{1}{4}\mu_0|H_y|^2 = \frac{1}{4}\epsilon_0 E_0^2\,\frac{n_1^2 n_3^2}{n_1^2 + n_3^2 - N^2} = w_e \\ w &= w_e + w_m = 2w_e = \frac{1}{2}\epsilon_0 E_0^2\,\frac{n_1^2 n_3^2}{n_1^2 + n_3^2 - N^2} \end{aligned} \qquad (4.6.30)$$

The vector $\boldsymbol{\mathcal{P}}$ is orthogonal to $\boldsymbol{E}$ and its direction is $\bar{\theta}$ given by Eq. (4.6.13), as can be verified by taking the ratio $\tan\bar{\theta} = \mathcal{P}_x/\mathcal{P}_z$. The energy transport velocity is the ratio of the energy flux to the energy density—it agrees with the group velocity (4.6.12):

$$\boldsymbol{v} = \frac{\boldsymbol{\mathcal{P}}}{w} = c_0\left(\hat{\mathbf{x}}\,\frac{N}{n_3^2}\sin\theta + \hat{\mathbf{z}}\,\frac{N}{n_1^2}\cos\theta\right) \qquad (4.6.31)$$

To summarize, the TE and TM uniform plane wave solutions are given by Eqs. (4.6.18) and (4.6.27). We will use these results in Sects. 8.10 and 8.12 to discuss reflection and refraction in birefringent media and multilayer birefringent dielectric structures. Further discussion of propagation in birefringent media can be found in [598] and [658–678].

4.7 Problems

4.1 For the circular-polarization basis of Eq. (4.1.1), show

$$\mathbf{E} = \hat{\mathbf{e}}_+ E_+ + \hat{\mathbf{e}}_- E_- \quad\Rightarrow\quad \hat{\mathbf{z}}\times\mathbf{E} = j\,\hat{\mathbf{e}}_+ E_+ - j\,\hat{\mathbf{e}}_- E_- \quad\Rightarrow\quad \hat{\mathbf{z}}\times\mathbf{E}_\pm = \pm j\mathbf{E}_\pm$$

4.2 Show the component-wise Maxwell equations, Eqs. (4.1.4) and (4.1.5), with respect to the linear and circular polarization bases.

4.3 Suppose that the two unit vectors $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}\}$ are rotated about the z-axis by an angle ϕ resulting in $\hat{\mathbf{x}}' = \hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi$ and $\hat{\mathbf{y}}' = \hat{\mathbf{y}}\cos\phi - \hat{\mathbf{x}}\sin\phi$. Show that the corresponding circular basis vectors $\hat{\mathbf{e}}_\pm = \hat{\mathbf{x}} \mp j\,\hat{\mathbf{y}}$ and $\hat{\mathbf{e}}'_\pm = \hat{\mathbf{x}}' \mp j\,\hat{\mathbf{y}}'$ change by the phase factors: $\hat{\mathbf{e}}'_\pm = e^{\pm j\phi}\hat{\mathbf{e}}_\pm$.

4.4 Consider a linearly birefringent 90° quarter-wave retarder. Show that the following input polarizations change into the indicated output ones:

$$\begin{aligned} \hat{\mathbf{x}} \pm \hat{\mathbf{y}} &\;\to\; \hat{\mathbf{x}} \pm j\,\hat{\mathbf{y}} \\ \hat{\mathbf{x}} \pm j\,\hat{\mathbf{y}} &\;\to\; \hat{\mathbf{x}} \pm \hat{\mathbf{y}} \end{aligned}$$

What are the output polarizations if the same input polarizations go through a 180° half-wave retarder?

4.5 A polarizer lets through linearly polarized light in the direction of the unit vector $\hat{\mathbf{e}}_p = \hat{\mathbf{x}}\cos\theta_p + \hat{\mathbf{y}}\sin\theta_p$, as shown in Fig. 4.7.1. The output of the polarizer propagates in the z-direction through a linearly birefringent retarder of length l, with birefringent refractive indices n_1, n_2, and retardance $\phi = (n_1 - n_2)k_0 l$.

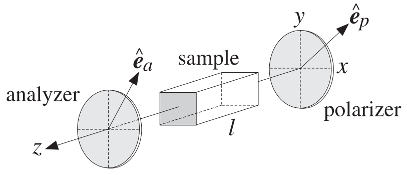


Fig. 4.7.1 Polarizer-analyzer measurement of birefringence.

The output $\mathbf{E}(l)$ of the birefringent sample goes through an analyzing linear polarizer that lets through polarizations along the unit vector $\hat{\mathbf{e}}_a = \hat{\mathbf{x}}\cos\theta_a + \hat{\mathbf{y}}\sin\theta_a$. Show that the light intensity at the output of the analyzer is given by:

$$I_a = \left|\hat{\mathbf{e}}_a\cdot\mathbf{E}(l)\right|^2 = \left|\cos\theta_a\cos\theta_p + e^{j\phi}\sin\theta_a\sin\theta_p\right|^2$$

For a circularly birefringent sample that introduces a natural or Faraday rotation of $\phi = (k_+ - k_-)l/2$, show that the output light intensity will be:

$$I_a = \left|\hat{\mathbf{e}}_a\cdot\mathbf{E}(l)\right|^2 = \cos^2(\theta_p - \theta_a - \phi)$$

For both the linear and circular cases, what are some convenient choices for θ_a and θ_p?

4.6 A linearly polarized wave with polarization direction at an angle θ with the x-axis goes through a circularly birefringent retarder that introduces an optical rotation by the angle $\phi = (k_+ - k_-)l/2$. Show that the input and output polarization directions will be:

$$\hat{\mathbf{x}}\cos\theta + \hat{\mathbf{y}}\sin\theta \;\to\; \hat{\mathbf{x}}\cos(\theta-\phi) + \hat{\mathbf{y}}\sin(\theta-\phi)$$

4.7 Show that an arbitrary polarization vector can be expressed as follows with respect to a linear basis $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}\}$ and its rotated version $\{\hat{\mathbf{x}}', \hat{\mathbf{y}}'\}$:

$$\mathbf{E} = A\,\hat{\mathbf{x}} + B\,\hat{\mathbf{y}} = A'\,\hat{\mathbf{x}}' + B'\,\hat{\mathbf{y}}'$$

where the new coefficients and the new basis vectors are related to the old ones by a rotation by an angle ϕ:

$$\begin{bmatrix} A' \\ B' \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix}, \qquad \begin{bmatrix} \hat{\mathbf{x}}' \\ \hat{\mathbf{y}}' \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}\begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{bmatrix}$$

4.8 Show that the source-free Maxwell's equations (4.1.2) for a chiral medium characterized by (4.3.1), may be cast in the matrix form, where $k = \omega\sqrt{\mu\epsilon}$, $\eta = \sqrt{\mu/\epsilon}$, and $a = \chi/\sqrt{\mu\epsilon}$:

$$\nabla\times\begin{bmatrix} \mathbf{E} \\ \eta\mathbf{H} \end{bmatrix} = \begin{bmatrix} ka & -jk \\ jk & ka \end{bmatrix}\begin{bmatrix} \mathbf{E} \\ \eta\mathbf{H} \end{bmatrix}$$

Show that these may be decoupled by forming the "right" and "left" polarized fields:

$$\nabla\times\begin{bmatrix} \mathbf{E}_R \\ \mathbf{E}_L \end{bmatrix} = \begin{bmatrix} k_+ & 0 \\ 0 & -k_- \end{bmatrix}\begin{bmatrix} \mathbf{E}_R \\ \mathbf{E}_L \end{bmatrix}, \quad \text{where} \quad \mathbf{E}_R = \frac{1}{2}(\mathbf{E} - j\eta\mathbf{H}), \quad \mathbf{E}_L = \frac{1}{2}(\mathbf{E} + j\eta\mathbf{H})$$

where $k_\pm = k(1 \pm a)$. Using these results, show that the possible plane-wave solutions propagating in the direction of a unit-vector $\hat{\mathbf{k}}$ are given by:

$$\mathbf{E}(r) = E_0(\hat{\mathbf{p}} - j\,\hat{\mathbf{s}})e^{-j\mathbf{k}_+\cdot\mathbf{r}} \quad\text{and}\quad \mathbf{E}(r) = E_0(\hat{\mathbf{p}} + j\,\hat{\mathbf{s}})e^{-j\mathbf{k}_-\cdot\mathbf{r}}$$

where $\mathbf{k}_\pm = k_\pm\hat{\mathbf{k}}$ and $\{\hat{\mathbf{p}}, \hat{\mathbf{s}}, \hat{\mathbf{k}}\}$ form a right-handed system of unit vectors, such as $\{\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}'\}$ of Fig. 2.9.1. Determine expressions for the corresponding magnetic fields. What freedom do we have in selecting $\{\hat{\mathbf{p}}, \hat{\mathbf{s}}\}$ for a given direction $\hat{\mathbf{k}}$?

4.9 Using Maxwell's equations (4.1.2), show the following Poynting-vector relationships for an arbitrary source-free medium:

$$\begin{aligned} \nabla\cdot(\mathbf{E}\times\mathbf{H}^*) &= j\omega(\mathbf{D}^*\cdot\mathbf{E} - \mathbf{B}\cdot\mathbf{H}^*) \\ \nabla\cdot\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*) &= -\omega\,\mathrm{Im}(\mathbf{D}^*\cdot\mathbf{E} + \mathbf{B}^*\cdot\mathbf{H}) \end{aligned}$$

Explain why a lossless medium must satisfy the condition $\nabla\cdot\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*) = 0$. Show that this condition requires that the energy function $w = (\mathbf{D}^*\cdot\mathbf{E} + \mathbf{B}^*\cdot\mathbf{H})/2$ be real-valued.

For a lossless chiral medium characterized by (4.3.1), show that the parameters ϵ, μ, χ are required to be real. Moreover, show that the positivity of the energy function $w > 0$ requires that $|\chi| < \sqrt{\mu\epsilon}$, as well as $\epsilon > 0$ and $\mu > 0$.

4.10 In a chiral medium, at $z = 0$ we lauch the fields $E_{R+}(0)$ and $E_{L-}(0)$, which propagate by a distance l, get reflected, and come back to the starting point. Assume that at the point of reversal the fields remain unchanged, that is, $E_{R+}(l) = E_{L+}(l)$ and $E_{L-}(l) = E_{R-}(l)$. Using the propagation results (4.3.5) and (4.3.9), show that fields returned back at $z = 0$ will be:

$$\begin{aligned} E_{L+}(0) &= E_{L+}(l)e^{-jk_-l} = E_{R+}(l)e^{-jk_-l} = E_{R+}(0)e^{-j(k_++k_-)l} \\ E_{R-}(0) &= E_{R-}(l)e^{-jk_+l} = E_{L-}(l)e^{-jk_+l} = E_{L-}(0)e^{-j(k_++k_-)l} \end{aligned}$$

Show that the overall natural rotation angle will be zero. For a gyrotropic medium, show that the corresponding rountrip fields will be:

$$\begin{aligned} E_{L+}(0) &= E_{L+}(l)\,e^{-jk_-l} = E_{R+}(l)\,e^{-jk_-l} = E_{R+}(0)\,e^{-2jk_+l} \\ E_{R-}(0) &= E_{R-}(l)\,e^{-jk_+l} = E_{L-}(l)\,e^{-jk_+l} = E_{L-}(0)\,e^{-2jk_-l} \end{aligned}$$

Show that the total Faraday rotation angle will be $2\phi = (k_+ - k_-)l$.

4.11 Show that the x, y components of the gyroelectric and gyromagnetic constitutive relationships (4.4.1) and (4.4.2) may be written in the compact forms:

$$\begin{aligned} \boldsymbol{D}_T &= \epsilon_1 \boldsymbol{E}_T - j\epsilon_2\,\hat{\mathbf{z}}\times\boldsymbol{E}_T \qquad \text{(gyroelectric)} \\ \boldsymbol{B}_T &= \mu_1 \boldsymbol{H}_T - j\mu_2\,\hat{\mathbf{z}}\times\boldsymbol{H}_T \qquad \text{(gyromagnetic)} \end{aligned}$$

where the subscript T indicates the transverse (with respect to z) part of a vector, for example, $\boldsymbol{D}_T = \hat{\mathbf{x}}D_x + \hat{\mathbf{y}}D_y$.

4.12 Conductors and plasmas exhibit gyroelectric behavior when they are in the presence of an external magnetic field. The equation of motion of conduction electrons in a constant magnetic field is $m\dot{\boldsymbol{v}} = e(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}) - m\alpha\boldsymbol{v}$, with the collisional damping term included. The magnetic field is in the z-direction, $\boldsymbol{B} = \hat{\mathbf{z}}B_0$.

Assuming $e^{j\omega t}$ time dependence and decomposing all vectors in the circular basis (4.1.1), for example, $\boldsymbol{v} = \hat{\mathbf{e}}_+ v_+ + \hat{\mathbf{e}}_- v_- + \hat{\mathbf{z}}\,v_z$, show that the solution of the equation of motion is:

$$v_\pm = \frac{\dfrac{e}{m}E_\pm}{\alpha + j(\omega \pm \omega_B)}\,, \quad v_z = \frac{\dfrac{e}{m}E_z}{\alpha + j\omega}$$

where $\omega_B = eB_0/m$ is the cyclotron frequency. Then, show that the $D{-}E$ constitutive relationship takes the form of Eq. (4.4.1) with:

$$\epsilon_\pm = \epsilon_1 \pm \epsilon_2 = \epsilon_0\left[1 - \frac{j\omega_p^2}{\omega\big[\alpha + j(\omega\pm\omega_B)\big]}\right], \quad \epsilon_3 = \epsilon_0\left[1 - \frac{j\omega_p^2}{\omega(\alpha + j\omega)}\right]$$

where $\omega_p^2 = Ne^2/m\epsilon_0$ is the plasma frequency and N, the number of conduction electrons per unit volume. (See Problem 1.10 for some helpful hints.)

4.13 If the magnetic field $\boldsymbol{H}_{\text{tot}} = \hat{\mathbf{z}}H_0 + \boldsymbol{H}e^{j\omega t}$ is applied to a magnetizable sample, the induced magnetic moment per unit volume (the magnetization) will have the form $\boldsymbol{M}_{\text{tot}} = \hat{\mathbf{z}}M_0 + \boldsymbol{M}e^{j\omega t}$, where $\hat{\mathbf{z}}M_0$ is the saturation magnetization due to $\hat{\mathbf{z}}H_0$ acting alone. The phenomenological equations governing $\boldsymbol{M}_{\text{tot}}$, including a so-called Landau-Lifshitz damping term, are given by [709]:

$$\frac{d\boldsymbol{M}_{\text{tot}}}{dt} = \gamma(\boldsymbol{M}_{\text{tot}}\times\boldsymbol{H}_{\text{tot}}) - \frac{\alpha}{M_0H_0}\,\boldsymbol{M}_{\text{tot}}\times(\boldsymbol{M}_{\text{tot}}\times\boldsymbol{H}_{\text{tot}})$$

where γ is the gyromagnetic ratio and $\tau = 1/\alpha$, a relaxation time constant. Assuming that $|\boldsymbol{H}| \ll H_0$ and $|\boldsymbol{M}| \ll M_0$, show that the linearized version of this equation obtained by keeping only first order terms in $\boldsymbol{H}$ and $\boldsymbol{M}$ is:

$$j\omega\boldsymbol{M} = \omega_M(\hat{\mathbf{z}}\times\boldsymbol{H}) - \omega_H(\hat{\mathbf{z}}\times\boldsymbol{M}) - \alpha\,\hat{\mathbf{z}}\times\big[(\boldsymbol{M} - \chi_0\boldsymbol{H})\times\hat{\mathbf{z}}\big]$$

where $\omega_M = \gamma M_0$, $\omega_H = \gamma H_0$, and $\chi_0 = M_0/H_0$. Working in the circular basis (4.1.1), show that the solution of this equation is:

$$M_\pm = \chi_0\,\frac{\alpha \pm j\omega_H}{\alpha + j(\omega\pm\omega_H)}\,H_\pm \equiv \chi_\pm H_\pm \quad \text{and} \quad M_z = 0$$

Writing $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$, show that the permeability matrix has the gyromagnetic form of Eq. (4.4.2) with $\mu_1 \pm \mu_2 = \mu_\pm = \mu_0(1 + \chi_\pm)$ and $\mu_3 = \mu_0$. Show that the real and imaginary parts of μ_1 are given by [709]:

$$\begin{aligned} \text{Re}(\mu_1) &= \mu_0 + \frac{\mu_0\chi_0}{2}\left[\frac{\alpha^2 + \omega_H(\omega+\omega_H)}{\alpha^2 + (\omega+\omega_H)^2} + \frac{\alpha^2 - \omega_H(\omega-\omega_H)}{\alpha^2 + (\omega-\omega_H)^2}\right] \\ \text{Im}(\mu_1) &= -\frac{\mu_0\chi_0}{2}\left[\frac{\alpha\omega}{\alpha^2 + (\omega+\omega_H)^2} + \frac{\alpha\omega}{\alpha^2 + (\omega-\omega_H)^2}\right] \end{aligned}$$

Derive similar expressions for $\text{Re}(\mu_2)$ and $\text{Im}(\mu_2)$.

4.14 A uniform plane wave, $\boldsymbol{E}e^{-j\boldsymbol{k}\cdot\mathbf{r}}$ and $\boldsymbol{H}e^{-j\boldsymbol{k}\cdot\mathbf{r}}$, is propagating in the direction of the unit vector $\hat{\mathbf{k}} = \hat{\mathbf{z}}' = \hat{\mathbf{z}}\cos\theta + \hat{\mathbf{z}}\sin\theta$ shown in Fig. 2.9.1 in a gyroelectric medium with constitutive relationships (4.4.1).

Show that Eqs. (4.6.14)–(4.6.16) remain valid provided we define the effective refractive index N through the wavevector $\boldsymbol{k} = k\,\hat{\mathbf{k}}$, where $k = Nk_0$, $k_0 = \omega\sqrt{\mu\epsilon_0}$.

Working in the circular-polarization basis (4.1.1), that is, $\boldsymbol{E} = \hat{\mathbf{e}}_+E_+ + \hat{\mathbf{e}}_-E_- + \hat{\mathbf{z}}\,E_z$, where $E_\pm = (E_x \pm jE_y)/2$, show that Eq. (4.6.16) leads to the homogeneous system:

$$\begin{bmatrix} 1 - \frac{1}{2}\sin^2\theta - \frac{\epsilon_+}{\epsilon_0N^2} & -\frac{1}{2}\sin^2\theta & -\frac{1}{2}\sin\theta\cos\theta \\ -\frac{1}{2}\sin^2\theta & 1 - \frac{1}{2}\sin^2\theta - \frac{\epsilon_-}{\epsilon_0N^2} & -\frac{1}{2}\sin\theta\cos\theta \\ -\sin\theta\cos\theta & -\sin\theta\cos\theta & \sin^2\theta - \frac{\epsilon_3}{\epsilon_0N^2} \end{bmatrix} \begin{bmatrix} E_+ \\ E_- \\ E_z \end{bmatrix} = 0 \tag{4.7.1}$$

where $\epsilon_\pm = \epsilon_1 \pm \epsilon_2$. Alternatively, show that in the linear-polarization basis:

$$\begin{bmatrix} \epsilon_1 - \epsilon_0N^2\cos^2\theta & j\epsilon_2 & \epsilon_0N^2\sin\theta\cos\theta \\ -j\epsilon_2 & \epsilon_1 - \epsilon_0N^2 & 0 \\ \epsilon_0N^2\sin\theta\cos\theta & 0 & \epsilon_3 - \epsilon_0N^2\sin^2\theta \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = 0 \tag{4.7.2}$$

For either basis, setting the determinant of the coefficient matrix to zero, show that a nonzero $\boldsymbol{E}$ solution exists provided that N^2 is one of the two solutions of:

$$\tan^2\theta = -\frac{\epsilon_3}{\epsilon_1}\,\frac{(\epsilon_0N^2 - \epsilon_+)(\epsilon_0N^2 - \epsilon_-)}{(\epsilon_0N^2 - \epsilon_3)(\epsilon_0N^2 - \epsilon_e)}\,, \quad \text{where} \quad \epsilon_e = \frac{2\epsilon_+\epsilon_-}{\epsilon_+ + \epsilon_-} = \frac{\epsilon_1^2 - \epsilon_2^2}{\epsilon_1} \tag{4.7.3}$$

Show that the two solutions for N^2 are:

$$N^2 = \frac{(\epsilon_1^2 - \epsilon_2^2 - \epsilon_1\epsilon_3)\sin^2\theta + 2\epsilon_1\epsilon_3 \pm \sqrt{(\epsilon_1^2 - \epsilon_2^2 - \epsilon_1\epsilon_3)^2\sin^4\theta + 4\epsilon_2^2\epsilon_3^2\cos^2\theta}}{2\epsilon_0(\epsilon_1\sin^2\theta + \epsilon_3\cos^2\theta)} \tag{4.7.4}$$

For the special case $\hat{\mathbf{k}} = \hat{\mathbf{z}}$ ($\theta = 0^{\text{o}}$), show that the two possible solutions of Eq. (4.7.1) are:

$$\begin{aligned} \epsilon_0N^2 &= \epsilon_+\,, \quad k = k_+ = \omega\sqrt{\mu\epsilon_+}\,, \quad E_+ \neq 0, \quad E_- = 0, \quad E_z = 0 \\ \epsilon_0N^2 &= \epsilon_-\,, \quad k = k_- = \omega\sqrt{\mu\epsilon_-}\,, \quad E_+ = 0, \quad E_- \neq 0, \quad E_z = 0 \end{aligned}$$

For the case $\hat{\mathbf{k}} = \hat{\mathbf{x}}$ ($\theta = 90^{\text{o}}$), show that:

$$\begin{aligned}
&\epsilon_0 N^2 = \epsilon_3\,, \quad k = k_3 = \omega\sqrt{\mu\epsilon_3}\,, \quad E_+ = 0, \quad E_- = 0, \qquad\qquad E_z \neq 0 \\
&\epsilon_0 N^2 = \epsilon_e\,, \quad k = k_e = \omega\sqrt{\mu\epsilon_e}\,, \quad E_+ \neq 0, \quad E_- = -\frac{\epsilon_+}{\epsilon_-}E_+\,, \quad E_z = 0
\end{aligned}$$

For each of the above four special solutions, derive the corresponding magnetic fields $\boldsymbol{H}$. Justify the four values of N^2 on the basis of Eq. (4.7.3). Discuss the polarization properties of the four cases. For the "extraordinary" wave $k = k_e$, show that $D_x = 0$ and $E_x/E_y = -j\epsilon_2/\epsilon_1$. Eq. (4.7.4) and the results of Problem 4.14 lead to the so-called Appleton-Hartree equations for describing plasma waves in a magnetic field [696–700].

4.15 A uniform plane wave, $\boldsymbol{E}e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}$ and $\boldsymbol{H}e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}$, is propagating in the direction of the unit vector $\hat{\mathbf{k}} = \hat{\mathbf{z}}' = \hat{\mathbf{z}}\cos\theta + \hat{\mathbf{z}}\sin\theta$ shown in Fig. 2.9.1 in a gyromagnetic medium with constitutive relationships (4.4.2). Using Maxwell's equations, show that:

$$\begin{aligned}
\boldsymbol{k}\times\boldsymbol{E} &= \omega\boldsymbol{B}\,, &\quad \boldsymbol{k}\cdot\boldsymbol{B} &= 0 \\
\boldsymbol{k}\times\boldsymbol{H} &= -\omega\epsilon\boldsymbol{E}\,, &\quad \boldsymbol{k}\cdot\boldsymbol{E} &= 0
\end{aligned} \quad\Rightarrow\quad \boldsymbol{H} - \frac{1}{\mu_0 N^2}\,\boldsymbol{B} = \hat{\mathbf{k}}(\hat{\mathbf{k}}\cdot\boldsymbol{H}) \tag{4.7.5}$$

where the effective refractive index N is defined through the wavevector $\boldsymbol{k} = k\,\hat{\mathbf{k}}$, where $k = Nk_0$, $k_0 = \omega\sqrt{\mu_0\epsilon}$. Working in the circular polarization basis $\boldsymbol{H} = \hat{\mathbf{e}}_+ H_+ + \hat{\mathbf{e}}_- H_- + \hat{\mathbf{z}}\,H_z$, where $H_\pm = (H_x \pm jH_y)/2$, show that Eq. (4.7.5) leads to the homogeneous system:

$$\begin{bmatrix}
1 - \frac{1}{2}\sin^2\theta - \frac{\mu_+}{\mu_0 N^2} & -\frac{1}{2}\sin^2\theta & -\frac{1}{2}\sin\theta\cos\theta \\
-\frac{1}{2}\sin^2\theta & 1 - \frac{1}{2}\sin^2\theta - \frac{\mu_-}{\mu_0 N^2} & -\frac{1}{2}\sin\theta\cos\theta \\
-\sin\theta\cos\theta & -\sin\theta\cos\theta & \sin^2\theta - \frac{\mu_3}{\mu_0 N^2}
\end{bmatrix}
\begin{bmatrix} H_+ \\ H_- \\ H_z \end{bmatrix} = 0 \tag{4.7.6}$$

where $\mu_\pm = \mu_1 \pm \mu_2$. Alternatively, show that in the linear-polarization basis:

$$\begin{bmatrix}
\mu_1 - \mu_0 N^2\cos^2\theta & j\mu_2 & \mu_0 N^2\sin\theta\cos\theta \\
-j\mu_2 & \mu_1 - \mu_0 N^2 & 0 \\
\mu_0 N^2\sin\theta\cos\theta & 0 & \mu_3 - \mu_0 N^2\sin^2\theta
\end{bmatrix}
\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = 0 \tag{4.7.7}$$

For either basis, setting the determinant of the coefficient matrix to zero, show that a nonzero $\boldsymbol{E}$ solution exists provided that N^2 is one of the two solutions of:

$$\tan^2\theta = -\frac{\mu_3}{\mu_1}\,\frac{(\mu_0 N^2 - \mu_+)(\mu_0 N^2 - \mu_-)}{(\mu_0 N^2 - \mu_3)(\mu_0 N^2 - \mu_e)}\,, \quad \text{where} \quad \mu_e = \frac{2\mu_+\mu_-}{\mu_+ + \mu_-} = \frac{\mu_1^2 - \mu_2^2}{\mu_1} \tag{4.7.8}$$

Show that the two solutions for N^2 are:

$$N^2 = \frac{(\mu_1^2 - \mu_2^2 - \mu_1\mu_3)\sin^2\theta + 2\mu_1\mu_3 \pm \sqrt{(\mu_1^2 - \mu_2^2 - \mu_1\mu_3)^2\sin^4\theta + 4\mu_2^2\mu_3^2\cos^2\theta}}{2\mu_0(\mu_1\sin^2\theta + \mu_3\cos^2\theta)}$$

For the special case $\theta = 0^{\text{o}}$, show that the two possible solutions of Eq. (4.7.6) are:

$$\begin{aligned}
&\mu_0 N^2 = \mu_+\,, \quad k = k_+ = \omega\sqrt{\epsilon\mu_+}\,, \quad H_+ \neq 0, \quad H_- = 0, \quad H_z = 0 \\
&\mu_0 N^2 = \mu_-\,, \quad k = k_+ = \omega\sqrt{\epsilon\mu_-}\,, \quad H_+ = 0, \quad H_- \neq 0, \quad H_z = 0
\end{aligned}$$

For the special case $\theta = 90^{\text{o}}$, show that:

$$\begin{aligned}
&\mu_0 N^2 = \mu_3\,, \quad k = k_3 = \omega\sqrt{\epsilon\mu_3}\,, \quad H_+ = 0, \quad H_- = 0, \qquad\qquad H_z \neq 0 \\
&\mu_0 N^2 = \mu_e\,, \quad k = k_e = \omega\sqrt{\epsilon\mu_e}\,, \quad H_+ \neq 0, \quad H_- = -\frac{\mu_+}{\mu_-}H_+\,, \quad H_z = 0
\end{aligned}$$

For each of the above four special solutions, derive the corresponding electric fields $\boldsymbol{E}$. Justify the four values of N^2 on the basis of Eq. (4.7.8). Discuss the polarization properties of the four cases. This problem is the dual of Problem 4.14.

4.16 Using Eq. (4.6.9) for the effective TM refractive index in a birefringent medium, show the following additional relationships:

$$\frac{\sin^2\theta}{1 - \dfrac{n_1^2}{N^2}} + \frac{\cos^2\theta}{1 - \dfrac{n_3^2}{N^2}} = 1 \tag{4.7.9}$$

$$\frac{n_3}{n_1}\cos^2\theta + \frac{n_1}{n_3}\sin^2\theta = \frac{n_1 n_3}{N^2} \tag{4.7.10}$$

$$\frac{n_1^2}{n_3^2}\sin^2\theta + \frac{n_3^2}{n_1^2}\cos^2\theta = \frac{n_1^2 + n_3^2 - N^2}{N^2} \tag{4.7.11}$$

$$\sin^2\theta = \frac{1 - \dfrac{n_1^2}{N^2}}{1 - \dfrac{n_1^2}{n_3^2}}\,, \qquad \cos^2\theta = \frac{1 - \dfrac{n_3^2}{N^2}}{1 - \dfrac{n_3^2}{n_1^2}} \tag{4.7.12}$$

$$\cos^2\theta - \frac{n_1^2}{N^2} = -\frac{n_1^2}{n_3^2}\sin^2\theta\,, \qquad \sin^2\theta - \frac{n_3^2}{N^2} = -\frac{n_3^2}{n_1^2}\cos^2\theta \tag{4.7.13}$$

Using these relationships, show that the homogeneous linear system (4.6.20) can be simplified into the form:

$$E_x\frac{n_1}{n_3}\sin\theta = -E_z\frac{n_3}{n_1}\cos\theta\,, \quad E_z\frac{n_3}{n_1}\cos\theta = -E_x\frac{n_1}{n_3}\sin\theta$$

5

Reflection and Transmission

5.1 Propagation Matrices

In this chapter, we consider uniform plane waves incident *normally* on material interfaces. Using the boundary conditions for the fields, we will relate the forward-backward fields on one side of the interface to those on the other side, expressing the relationship in terms of a 2×2 *matching matrix*.

If there are several interfaces, we will propagate our forward-backward fields from one interface to the next with the help of a 2×2 *propagation matrix*. The combination of a matching and a propagation matrix relating the fields across different interfaces will be referred to as a *transfer* or transition matrix.

We begin by discussing propagation matrices. Consider an electric field that is linearly polarized in the x-direction and propagating along the z-direction in a lossless (homogeneous and isotropic) dielectric. Setting $\boldsymbol{E}(z)=\hat{\mathbf{x}}E_x(z)=\hat{\mathbf{x}}E(z)$ and $\boldsymbol{H}(z)=\hat{\mathbf{y}}H_y(z)=\hat{\mathbf{y}}H(z)$, we have from Eq. (2.2.6):

$$\begin{aligned} E(z) &= E_{0+}e^{-jkz}+E_{0-}e^{jkz}=E_+(z)+E_-(z) \\ H(z) &= \frac{1}{\eta}\big[E_{0+}e^{-jkz}-E_{0-}e^{jkz}\big]=\frac{1}{\eta}\big[E_+(z)-E_-(z)\big] \end{aligned} \tag{5.1.1}$$

where the corresponding forward and backward electric fields at position z are:

$$\begin{aligned} E_+(z) &= E_{0+}e^{-jkz} \\ E_-(z) &= E_{0-}e^{jkz} \end{aligned} \tag{5.1.2}$$

We can also express the fields $E_\pm(z)$ in terms of $E(z),H(z)$. Adding and subtracting the two equations (5.1.1), we find:

$$\begin{aligned} E_+(z) &= \frac{1}{2}\big[E(z)+\eta H(z)\big] \\ E_-(z) &= \frac{1}{2}\big[E(z)-\eta H(z)\big] \end{aligned} \tag{5.1.3}$$

Eqs.(5.1.1) and (5.1.3) can also be written in the convenient matrix forms:

$$\begin{bmatrix} E \\ H \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \eta^{-1} & -\eta^{-1} \end{bmatrix}\begin{bmatrix} E_+ \\ E_- \end{bmatrix}, \qquad \begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & \eta \\ 1 & -\eta \end{bmatrix}\begin{bmatrix} E \\ H \end{bmatrix} \tag{5.1.4}$$

Two useful quantities in interface problems are the *wave impedance* at z:

$$\boxed{Z(z) = \frac{E(z)}{H(z)}} \qquad \text{(wave impedance)} \tag{5.1.5}$$

and the *reflection coefficient* at position z:

$$\boxed{\Gamma(z) = \frac{E_-(z)}{E_+(z)}} \qquad \text{(reflection coefficient)} \tag{5.1.6}$$

Using Eq. (5.1.3), we have:

$$\Gamma = \frac{E_-}{E_+} = \frac{\frac{1}{2}(E - \eta H)}{\frac{1}{2}(E + \eta H)} = \frac{\frac{E}{H} - \eta}{\frac{E}{H} + \eta} = \frac{Z - \eta}{Z + \eta}$$

Similarly, using Eq. (5.1.1) we find:

$$Z = \frac{E}{H} = \frac{E_+ + E_-}{\frac{1}{\eta}(E_+ - E_-)} = \eta\,\frac{1 + \frac{E_-}{E_+}}{1 - \frac{E_-}{E_+}} = \eta\,\frac{1 + \Gamma}{1 - \Gamma}$$

Thus, we have the relationships:

$$\boxed{Z(z) = \eta\,\frac{1 + \Gamma(z)}{1 - \Gamma(z)}} \quad \Leftrightarrow \quad \boxed{\Gamma(z) = \frac{Z(z) - \eta}{Z(z) + \eta}} \tag{5.1.7}$$

Using Eq. (5.1.2), we find:

$$\Gamma(z) = \frac{E_-(z)}{E_+(z)} = \frac{E_{0-}e^{jkz}}{E_{0+}e^{-jkz}} = \Gamma(0)\,e^{2jkz}$$

where $\Gamma(0) = E_{0-}/E_{0+}$ is the reflection coefficient at $z = 0$. Thus,

$$\Gamma(z) = \Gamma(0)\,e^{2jkz} \qquad \text{(propagation of } \Gamma\text{)} \tag{5.1.8}$$

Applying (5.1.7) at z and $z = 0$, we have:

$$\frac{Z(z) - \eta}{Z(z) + \eta} = \Gamma(z) = \Gamma(0)\,e^{2jkz} = \frac{Z(0) - \eta}{Z(0) + \eta}e^{2jkz}$$

This may be solved for $Z(z)$ in terms of $Z(0)$, giving after some algebra:

$$Z(z) = \eta\,\frac{Z(0) - j\eta\tan kz}{\eta - jZ(0)\tan kz} \qquad \text{(propagation of } Z\text{)} \tag{5.1.9}$$

The reason for introducing so many field quantities is that the three quantities $\{E_+(z), E_-(z), \Gamma(z)\}$ have simple propagation properties, whereas $\{E(z), H(z), Z(z)\}$ do not. On the other hand, $\{E(z), H(z), Z(z)\}$ match simply across interfaces, whereas $\{E_+(z), E_-(z), \Gamma(z)\}$ do not.

Eqs. (5.1.1) and (5.1.2) relate the field quantities at location z to the quantities at $z = 0$. In matching problems, it proves more convenient to be able to relate these quantities at two arbitrary locations.

Fig. 5.1.1 depicts the quantities $\{E(z), H(z), E_+(z), E_-(z), Z(z), \Gamma(z)\}$ at the two locations z_1 and z_2 separated by a distance $l = z_2 - z_1$. Using Eq. (5.1.2), we have for the forward field at these two positions:

$$E_{2+} = E_{0+}e^{-jkz_2}, \qquad E_{1+} = E_{0+}e^{-jkz_1} = E_{0+}e^{-jk(z_2 - l)} = e^{jkl}E_{2+}$$

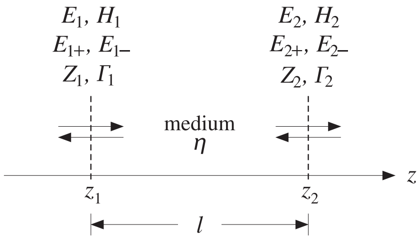


Fig. 5.1.1 Field quantities propagated between two positions in space.

And similarly, $E_{1-} = e^{-jkl}E_{2-}$. Thus,

$$E_{1+} = e^{jkl}E_{2+}, \qquad E_{1-} = e^{-jkl}E_{2-} \tag{5.1.10}$$

and in matrix form:

$$\begin{bmatrix} E_{1+} \\ E_{1-} \end{bmatrix} = \begin{bmatrix} e^{jkl} & 0 \\ 0 & e^{-jkl} \end{bmatrix}\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix} \qquad \text{(propagation matrix)} \tag{5.1.11}$$

We will refer to this as the *propagation matrix* for the forward and backward fields. It follows that the reflection coefficients will be related by:

$$\Gamma_1 = \frac{E_{1-}}{E_{1+}} = \frac{E_{2-}e^{-jkl}}{E_{2+}e^{jkl}} = \Gamma_2 e^{-2jkl}, \qquad \text{or,}$$

$$\boxed{\Gamma_1 = \Gamma_2 e^{-2jkl}} \qquad \text{(reflection coefficient propagation)} \tag{5.1.12}$$

Using the matrix relationships (5.1.4) and (5.1.11), we may also express the total electric and magnetic fields E_1, H_1 at position z_1 in terms of E_2, H_2 at position z_2:

$$\begin{aligned} \begin{bmatrix} E_1 \\ H_1 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ \eta^{-1} & -\eta^{-1} \end{bmatrix}\begin{bmatrix} E_{1+} \\ E_{1-} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \eta^{-1} & -\eta^{-1} \end{bmatrix}\begin{bmatrix} e^{jkl} & 0 \\ 0 & e^{-jkl} \end{bmatrix}\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ \eta^{-1} & -\eta^{-1} \end{bmatrix}\begin{bmatrix} e^{jkl} & 0 \\ 0 & e^{-jkl} \end{bmatrix}\begin{bmatrix} 1 & \eta \\ 1 & -\eta \end{bmatrix}\begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \end{aligned}$$

which gives after some algebra:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} \cos kl & j\eta\sin kl \\ j\eta^{-1}\sin kl & \cos kl \end{bmatrix} \begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \qquad \text{(propagation matrix)} \qquad (5.1.13)$$

Writing $\eta = \eta_0/n$, where n is the refractive index of the propagation medium, Eq. (5.1.13) can written in following form, which is useful in analyzing multilayer structures and is common in the thin-film literature [592,594,598,609]:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} \cos\delta & jn^{-1}\eta_0\sin\delta \\ jn\eta_0^{-1}\sin\delta & \cos\delta \end{bmatrix} \begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \qquad \text{(propagation matrix)} \qquad (5.1.14)$$

where δ is the propagation *phase constant*, $\delta = kl = k_0 nl = 2\pi(nl)/\lambda_0$, and nl the *optical length*. Eqs. (5.1.13) and (5.1.5), imply for the propagation of the wave impedance:

$$Z_1 = \frac{E_1}{H_1} = \frac{E_2\cos kl + j\eta H_2\sin kl}{jE_2\eta^{-1}\sin kl + H_2\cos kl} = \eta\,\frac{\dfrac{E_2}{H_2}\cos kl + j\eta\sin kl}{\eta\cos kl + j\dfrac{E_2}{H_2}\sin kl}$$

which gives:

$$\boxed{Z_1 = \eta\,\frac{Z_2\cos kl + j\eta\sin kl}{\eta\cos kl + jZ_2\sin kl}} \qquad \text{(impedance propagation)} \qquad (5.1.15)$$

It can also be written in the form:

$$\boxed{Z_1 = \eta\,\frac{Z_2 + j\eta\tan kl}{\eta + jZ_2\tan kl}} \qquad \text{(impedance propagation)} \qquad (5.1.16)$$

A useful way of expressing Z_1 is in terms of the reflection coefficient Γ_2. Using (5.1.7) and (5.1.12), we have:

$$Z_1 = \eta\,\frac{1+\Gamma_1}{1-\Gamma_1} = \eta\,\frac{1+\Gamma_2 e^{-2jkl}}{1-\Gamma_2 e^{-2jkl}} \qquad \text{or,}$$

$$\boxed{Z_1 = \eta\,\frac{1+\Gamma_2 e^{-2jkl}}{1-\Gamma_2 e^{-2jkl}}} \qquad (5.1.17)$$

We mention finally two special propagation cases: the *half-wavelength* and the *quarter-wavelength* cases. When the propagation distance is $l = \lambda/2$, or any integral multiple thereof, the wave impedance and reflection coefficient remain unchanged. Indeed, we have in this case $kl = 2\pi l/\lambda = 2\pi/2 = \pi$ and $2kl = 2\pi$. It follows from Eq. (5.1.12) that $\Gamma_1 = \Gamma_2$ and hence $Z_1 = Z_2$.

If on the other hand $l = \lambda/4$, or any odd integral multiple thereof, then $kl = 2\pi/4 = \pi/2$ and $2kl = \pi$. The reflection coefficient changes sign and the wave impedance inverts:

$$\Gamma_1 = \Gamma_2 e^{-2jkl} = \Gamma_2 e^{-j\pi} = -\Gamma_2 \quad \Rightarrow \quad Z_1 = \eta\,\frac{1+\Gamma_1}{1-\Gamma_1} = \eta\,\frac{1-\Gamma_2}{1+\Gamma_2} = \eta\,\frac{1}{Z_2/\eta} = \frac{\eta^2}{Z_2}$$

Thus, we have in the two cases:

$$\boxed{\begin{aligned} l &= \frac{\lambda}{2} \quad \Rightarrow \quad Z_1 = Z_2, \quad \Gamma_1 = \Gamma_2 \\ l &= \frac{\lambda}{4} \quad \Rightarrow \quad Z_1 = \frac{\eta^2}{Z_2}, \quad \Gamma_1 = -\Gamma_2 \end{aligned}} \qquad (5.1.18)$$

5.2 Matching Matrices

Next, we discuss the matching conditions across dielectric interfaces. We consider a planar interface (taken to be the xy-plane at some location z) separating two dielectric/conducting media with (possibly complex-valued) characteristic impedances η, η', as shown in Fig. 5.2.1.[†]

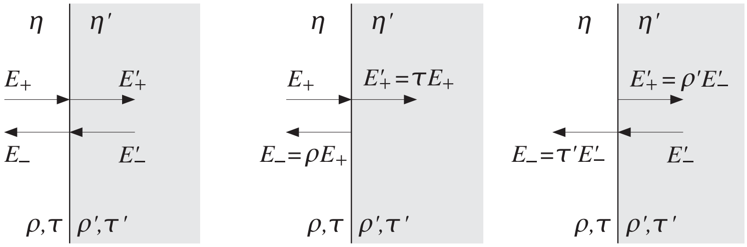


Fig. 5.2.1 Fields across an interface.

Because the normally incident fields are tangential to the interface plane, the boundary conditions require that the *total* electric and magnetic fields be continuous across the two sides of the interface:

$$\boxed{\begin{aligned} E &= E' \\ H &= H' \end{aligned}} \qquad \text{(continuity across interface)} \qquad (5.2.1)$$

In terms of the forward and backward electric fields, Eq. (5.2.1) reads:

$$\begin{aligned} E_+ + E_- &= E'_+ + E'_- \\ \frac{1}{\eta}(E_+ - E_-) &= \frac{1}{\eta'}(E'_+ - E'_-) \end{aligned} \qquad (5.2.2)$$

Eq. (5.2.2) may be written in a matrix form relating the fields $E_\pm$ on the left of the interface to the fields $E'_\pm$ on the right:

$$\begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} E'_+ \\ E'_- \end{bmatrix} \qquad \text{(matching matrix)} \qquad (5.2.3)$$

and inversely:

[†] The arrows in this figure indicate the directions of propagation, not the direction of the fields—the field vectors are perpendicular to the propagation directions and parallel to the interface plane.

$$\begin{bmatrix} E'_+ \\ E'_- \end{bmatrix} = \frac{1}{\tau'}\begin{bmatrix} 1 & \rho' \\ \rho' & 1 \end{bmatrix}\begin{bmatrix} E_+ \\ E_- \end{bmatrix} \qquad \text{(matching matrix)} \tag{5.2.4}$$

where $\{\rho, \tau\}$ and $\{\rho', \tau'\}$ are the elementary reflection and transmission coefficients from the left and from the right of the interface, defined in terms of η, η' as follows:

$$\boxed{\rho = \frac{\eta' - \eta}{\eta' + \eta}\,, \quad \tau = \frac{2\eta'}{\eta' + \eta}} \tag{5.2.5}$$

$$\boxed{\rho' = \frac{\eta - \eta'}{\eta + \eta'}\,, \quad \tau' = \frac{2\eta}{\eta + \eta'}} \tag{5.2.6}$$

Writing $\eta = \eta_0/n$ and $\eta' = \eta_0/n'$, we have in terms of the refractive indices:

$$\boxed{\begin{aligned} \rho &= \frac{n - n'}{n + n'}\,, & \tau &= \frac{2n}{n + n'} \\ \rho' &= \frac{n' - n}{n' + n}\,, & \tau' &= \frac{2n'}{n' + n} \end{aligned}} \tag{5.2.7}$$

These are also called the Fresnel coefficients. We note various useful relationships:

$$\tau = 1 + \rho, \quad \rho' = -\rho, \quad \tau' = 1 + \rho' = 1 - \rho, \quad \tau\tau' = 1 - \rho^2 \tag{5.2.8}$$

In summary, the total electric and magnetic fields E, H match simply across the interface, whereas the forward/backward fields $E_\pm$ are related by the matching matrices of Eqs. (5.2.3) and (5.2.4). An immediate consequence of Eq. (5.2.1) is that the wave impedance is *continuous* across the interface:

$$Z = \frac{E}{H} = \frac{E'}{H'} = Z'$$

On the other hand, the corresponding reflection coefficients $\Gamma = E_-/E_+$ and $\Gamma' = E'_-/E'_+$ match in a more complicated way. Using Eq. (5.1.7) and the continuity of the wave impedance, we have:

$$\eta\frac{1 + \Gamma}{1 - \Gamma} = Z = Z' = \eta'\frac{1 + \Gamma'}{1 - \Gamma'}$$

which can be solved to get:

$$\Gamma = \frac{\rho + \Gamma'}{1 + \rho\Gamma'} \quad \text{and} \quad \Gamma' = \frac{\rho' + \Gamma}{1 + \rho'\Gamma}$$

The same relationship follows also from Eq. (5.2.3):

$$\Gamma = \frac{E_-}{E_+} = \frac{\dfrac{1}{\tau}(\rho E'_+ + E'_-)}{\dfrac{1}{\tau}(E'_+ + \rho E'_-)} = \frac{\rho + \dfrac{E'_-}{E'_+}}{1 + \rho\dfrac{E'_-}{E'_+}} = \frac{\rho + \Gamma'}{1 + \rho\Gamma'}$$

To summarize, we have the matching conditions for Z and Γ:

$$\boxed{Z = Z'} \quad \Leftrightarrow \quad \boxed{\Gamma = \frac{\rho + \Gamma'}{1 + \rho\Gamma'}} \quad \Leftrightarrow \quad \boxed{\Gamma' = \frac{\rho' + \Gamma}{1 + \rho'\Gamma}} \tag{5.2.9}$$

Two special cases, illustrated in Fig. 5.2.1, are when there is only an incident wave on the interface from the left, so that $E'_- = 0$, and when the incident wave is only from the right, so that $E_+ = 0$. In the first case, we have $\Gamma' = E'_-/E'_+ = 0$, which implies $Z' = \eta'(1 + \Gamma')/(1 - \Gamma') = \eta'$. The matching conditions give then:

$$Z = Z' = \eta', \quad \Gamma = \frac{\rho + \Gamma'}{1 + \rho\Gamma'} = \rho$$

The matching matrix (5.2.3) implies in this case:

$$\begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} E'_+ \\ 0 \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} E'_+ \\ \rho E'_+ \end{bmatrix}$$

Expressing the reflected and transmitted fields E_-, E'_+ in terms of the incident field E_+, we have:

$$\boxed{\begin{aligned} E_- &= \rho E_+ \\ E'_+ &= \tau E_+ \end{aligned}} \qquad \text{(left-incident fields)} \tag{5.2.10}$$

This justifies the terms reflection and transmission coefficients for ρ and τ. In the right-incident case, the condition $E_+ = 0$ implies for Eq. (5.2.4):

$$\begin{bmatrix} E'_+ \\ E'_- \end{bmatrix} = \frac{1}{\tau'}\begin{bmatrix} 1 & \rho' \\ \rho' & 1 \end{bmatrix}\begin{bmatrix} 0 \\ E_- \end{bmatrix} = \frac{1}{\tau'}\begin{bmatrix} \rho' E_- \\ E_- \end{bmatrix}$$

These can be rewritten in the form:

$$\boxed{\begin{aligned} E'_+ &= \rho' E'_- \\ E_- &= \tau' E'_- \end{aligned}} \qquad \text{(right-incident fields)} \tag{5.2.11}$$

which relates the reflected and transmitted fields E'_+, E_- to the incident field E'_-. In this case $\Gamma = E_-/E_+ = \infty$ and the third of Eqs. (5.2.9) gives $\Gamma' = E'_-/E'_+ = 1/\rho'$, which is consistent with Eq. (5.2.11).

When there are incident fields both from both sides, that is, E_+, E'_-, we may invoke the linearity of Maxwell's equations and add the two right-hand sides of Eqs. (5.2.10) and (5.2.11) to obtain the outgoing fields E'_+, E_- in terms of the incident ones:

$$\boxed{\begin{aligned} E'_+ &= \tau E_+ + \rho' E'_- \\ E_- &= \rho E_+ + \tau' E'_- \end{aligned}} \tag{5.2.12}$$

This gives the *scattering matrix* relating the outgoing fields to the incoming ones:

$$\begin{bmatrix} E'_+ \\ E_- \end{bmatrix} = \begin{bmatrix} \tau & \rho' \\ \rho & \tau' \end{bmatrix}\begin{bmatrix} E_+ \\ E'_- \end{bmatrix} \qquad \text{(scattering matrix)} \tag{5.2.13}$$

Using the relationships Eq. (5.2.8), it is easily verified that Eq. (5.2.13) is equivalent to the matching matrix equations (5.2.3) and (5.2.4).

5.3 Reflected and Transmitted Power

For waves propagating in the z-direction, the time-averaged Poynting vector has only a z-component:

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\,\mathrm{Re}\big(\hat{\mathbf{x}}E\times\hat{\mathbf{y}}H^*\big) = \hat{\mathbf{z}}\,\frac{1}{2}\,\mathrm{Re}(EH^*)$$

A direct consequence of the continuity equations (5.2.1) is that the Poynting vector is conserved across the interface. Indeed, we have:

$$\mathcal{P} = \frac{1}{2}\,\mathrm{Re}(EH^*) = \frac{1}{2}\,\mathrm{Re}(E'H'^*) = \mathcal{P}' \tag{5.3.1}$$

In particular, consider the case of a wave incident from a lossless dielectric η onto a lossy dielectric η'. Then, the conservation equation (5.3.1) reads in terms of the forward and backward fields (assuming $E'_- = 0$):

$$\mathcal{P} = \frac{1}{2\eta}\big(|E_+|^2 - |E_-|^2\big) = \mathrm{Re}\Big(\frac{1}{2\eta'}\Big)|E'_+|^2 = \mathcal{P}'$$

The left hand-side is the difference of the incident and the reflected power and represents the amount of power transmitted into the lossy dielectric per unit area. We saw in Sec. 2.6 that this power is completely dissipated into heat inside the lossy dielectric (assuming it is infinite to the right.) Using Eqs. (5.2.10), we find:

$$\mathcal{P} = \frac{1}{2\eta}\,|E_+|^2\big(1-|\rho|^2\big) = \mathrm{Re}\Big(\frac{1}{2\eta'}\Big)|E_+|^2|\tau|^2 \tag{5.3.2}$$

This equality requires that:

$$\frac{1}{\eta}\big(1-|\rho|^2\big) = \mathrm{Re}\Big(\frac{1}{\eta'}\Big)|\tau|^2 \tag{5.3.3}$$

This can be proved using the definitions (5.2.5). Indeed, we have:

$$\frac{\eta}{\eta'} = \frac{1-\rho}{1+\rho} \quad\Rightarrow\quad \mathrm{Re}\left(\frac{\eta}{\eta'}\right) = \frac{1-|\rho|^2}{|1+\rho|^2} = \frac{1-|\rho|^2}{|\tau|^2}$$

which is equivalent to Eq. (5.3.3), if η is lossless (i.e., real.) Defining the incident, reflected, and transmitted powers by

$$\mathcal{P}_{\mathrm{in}} = \frac{1}{2\eta}|E_+|^2$$

$$\mathcal{P}_{\mathrm{ref}} = \frac{1}{2\eta}|E_-|^2 = \frac{1}{2\eta}|E_+|^2|\rho|^2 = \mathcal{P}_{\mathrm{in}}|\rho|^2$$

$$\mathcal{P}_{\mathrm{tr}} = \mathrm{Re}\Big(\frac{1}{2\eta'}\Big)|E'_+|^2 = \mathrm{Re}\Big(\frac{1}{2\eta'}\Big)|E_+|^2|\tau|^2 = \mathcal{P}_{\mathrm{in}}\,\mathrm{Re}\Big(\frac{\eta}{\eta'}\Big)|\tau|^2$$

Then, Eq. (5.3.2) reads $\mathcal{P}_{\mathrm{tr}} = \mathcal{P}_{\mathrm{in}} - \mathcal{P}_{\mathrm{ref}}$. The *power* reflection and transmission coefficients, also known as the *reflectance* and *transmittance*, give the percentage of the incident power that gets reflected and transmitted:

$$\frac{\mathcal{P}_{\mathrm{ref}}}{\mathcal{P}_{\mathrm{in}}} = |\rho|^2, \qquad \frac{\mathcal{P}_{\mathrm{tr}}}{\mathcal{P}_{\mathrm{in}}} = 1-|\rho|^2 = \mathrm{Re}\Big(\frac{\eta}{\eta'}\Big)|\tau|^2 = \mathrm{Re}\Big(\frac{n'}{n}\Big)|\tau|^2 \tag{5.3.4}$$

If both dielectrics are lossless, then ρ,τ are real-valued. In this case, if there are incident waves from both sides of the interface, it is straightforward to show that the net power moving towards the z-direction is the same at either side of the interface:

$$\mathcal{P} = \frac{1}{2\eta}\big(|E_+|^2 - |E_-|^2\big) = \frac{1}{2\eta'}\big(|E'_+|^2 - |E'_-|^2\big) = \mathcal{P}' \tag{5.3.5}$$

This follows from the matrix identity satisfied by the matching matrix of Eq. (5.2.3):

$$\frac{1}{\tau^2}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} = \frac{\eta}{\eta'}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \tag{5.3.6}$$

If ρ,τ are real, then we have with the help of this identity and Eq. (5.2.3):

$$\begin{aligned}
\mathcal{P} &= \frac{1}{2\eta}\big(|E_+|^2 - |E_-|^2\big) = \frac{1}{2\eta}\big[E_+^*, E_-^*\big]\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}E_+\\ E_-\end{bmatrix}\\
&= \frac{1}{2\eta}\big[E_+'^{\,*}, E_-'^{\,*}\big]\frac{1}{\tau\tau^*}\begin{bmatrix}1 & \rho^*\\ \rho^* & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}E'_+\\ E'_-\end{bmatrix}\\
&= \frac{1}{2\eta}\frac{\eta}{\eta'}\big[E_+'^{\,*}, E_-'^{\,*}\big]\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}E'_+\\ E'_-\end{bmatrix} = \frac{1}{2\eta'}\big(|E'_+|^2 - |E'_-|^2\big) = \mathcal{P}'
\end{aligned}$$

Example 5.3.1: Glasses have a refractive index of the order of $n = 1.5$ and dielectric constant $\epsilon = n^2\epsilon_0 = 2.25\epsilon_0$. Calculate the percentages of reflected and transmitted powers for visible light incident on a planar glass interface from air.

Solution: The characteristic impedance of glass will be $\eta = \eta_0/n$. Therefore, the reflection and transmission coefficients can be expressed directly in terms of n, as follows:

$$\rho = \frac{\eta-\eta_0}{\eta+\eta_0} = \frac{n^{-1}-1}{n^{-1}+1} = \frac{1-n}{1+n}, \qquad \tau = 1+\rho = \frac{2}{1+n}$$

For $n = 1.5$, we find $\rho = -0.2$ and $\tau = 0.8$. It follows that the power reflection and transmission coefficients will be

$$|\rho|^2 = 0.04, \qquad 1-|\rho|^2 = 0.96$$

That is, 4% of the incident power is reflected and 96% transmitted. □

Example 5.3.2: A uniform plane wave of frequency f is normally incident from air onto a thick conducting sheet with conductivity σ, and $\epsilon = \epsilon_0$, $\mu = \mu_0$. Show that the proportion of power transmitted into the conductor (and then dissipated into heat) is given approximately by

$$\frac{\mathcal{P}_{\mathrm{tr}}}{\mathcal{P}_{\mathrm{in}}} = \frac{4R_s}{\eta_0} = \sqrt{\frac{8\omega\epsilon_0}{\sigma}}$$

Calculate this quantity for $f = 1$ GHz and copper $\sigma = 5.8{\times}10^7$ Siemens/m.

Solution: For a good conductor, we have $\sqrt{\omega\epsilon_0/\sigma} \ll 1$. It follows from Eq. (2.8.4) that $R_s/\eta_0 = \sqrt{\omega\epsilon_0/2\sigma} \ll 1$. From Eq. (2.8.2), the conductor's characteristic impedance is $\eta_c = R_s(1+j)$. Thus, the quantity $\eta_c/\eta_0 = (1+j)R_s/\eta_0$ is also small. The reflection and transmission coefficients ρ, τ can be expressed to first-order in the quantity η_c/η_0 as follows:

$$\tau = \frac{2\eta_c}{\eta_c + \eta_0} \simeq \frac{2\eta_c}{\eta_0}, \qquad \rho = \tau - 1 \simeq -1 + \frac{2\eta_c}{\eta_0}$$

Similarly, the power transmission coefficient can be approximated as

$$1 - |\rho|^2 = 1 - |\tau - 1|^2 = 1 - 1 - |\tau|^2 + 2\,\mathrm{Re}(\tau) \simeq 2\,\mathrm{Re}(\tau) = 2\frac{2\,\mathrm{Re}(\eta_c)}{\eta_0} = \frac{4R_s}{\eta_0}$$

where we neglected $|\tau|^2$ as it is second order in η_c/η_0. For copper at 1 GHz, we have $\sqrt{\omega\epsilon_0/2\sigma} = 2.19\times10^{-5}$, which gives $R_s = \eta_0\sqrt{\omega\epsilon_0/2\sigma} = 377\times2.19\times10^{-5} = 0.0082\ \Omega$. It follows that $1 - |\rho|^2 = 4R_2/\eta_0 = 8.76\times10^{-5}$.

This represents only a small power loss of 8.76×10^{-3} percent and the sheet acts as very good mirror at microwave frequencies.

On the other hand, at optical frequencies, e.g., $f = 600$ THz corresponding to green light with $\lambda = 500$ nm, the exact equations (2.6.5) yield the value for the characteristic impedance of the sheet $\eta_c = 6.3924 + 6.3888i\ \Omega$ and the reflection coefficient $\rho = -0.9661 + 0.0328i$. The corresponding power loss is $1 - |\rho|^2 = 0.065$, or 6.5 percent. Thus, metallic mirrors are fairly lossy at optical frequencies. □

Example 5.3.3: A uniform plane wave of frequency f is normally incident from air onto a thick conductor with conductivity σ, and $\epsilon = \epsilon_0$, $\mu = \mu_0$. Determine the reflected and transmitted electric and magnetic fields to first-order in η_c/η_0 and in the limit of a perfect conductor ($\eta_c = 0$).

Solution: Using the approximations for ρ and τ of the previous example and Eq. (5.2.10), we have for the reflected, transmitted, and total electric fields at the interface:

$$\begin{aligned}
E_- &= \rho E_+ = \left(-1 + \frac{2\eta_c}{\eta_0}\right) E_+ \\
E'_+ &= \tau E_+ = \frac{2\eta_c}{\eta_0} E_+ \\
E &= E_+ + E_- = \frac{2\eta_c}{\eta_0} E_+ = E'_+ = E'
\end{aligned}$$

For a perfect conductor, we have $\sigma \to \infty$ and $\eta_c/\eta_0 \to 0$. The corresponding total tangential electric field becomes zero $E = E' = 0$, and $\rho = -1$, $\tau = 0$. For the magnetic fields, we need to develop similar first-order approximations. The incident magnetic field intensity is $H_+ = E_+/\eta_0$. The reflected field becomes to first order:

$$H_- = -\frac{1}{\eta_0}E_- = -\frac{1}{\eta_0}\rho E_+ = -\rho H_+ = \left(1 - \frac{2\eta_c}{\eta_0}\right) H_+$$

Similarly, the transmitted field is

$$H'_+ = \frac{1}{\eta_c}E'_+ = \frac{1}{\eta_c}\tau E_+ = \frac{\eta_0}{\eta_c}\tau H_+ = \frac{\eta_0}{\eta_c}\frac{2\eta_c}{\eta_c + \eta_0} H_+ = \frac{2\eta_0}{\eta_c + \eta_0} H_+ \simeq 2\left(1 - \frac{\eta_c}{\eta_0}\right) H_+$$

The total tangential field at the interface will be:

$$H = H_+ + H_- = 2\left(1 - \frac{\eta_c}{\eta_0}\right) H_+ = H'_+ = H'$$

In the perfect conductor limit, we find $H = H' = 2H_+$. As we saw in Sec. 2.6, the fields just inside the conductor, E'_+, H'_+, will attenuate while they propagate. Assuming the interface is at $z = 0$, we have:

$$E'_+(z) = E'_+ e^{-\alpha z} e^{-j\beta z}, \qquad H'_+(z) = H'_+ e^{-\alpha z} e^{-j\beta z}$$

where $\alpha = \beta = (1-j)/\delta$, and δ is the skin depth $\delta = \sqrt{\omega\mu\sigma/2}$. We saw in Sec. 2.6 that the effective surface current is equal in magnitude to the magnetic field at $z = 0$, that is, $J_s = H'_+$. Because of the boundary condition $H = H' = H'_+$, we obtain the result $J_s = H$, or vectorially, $\mathbf{J}_s = \mathbf{H} \times \hat{\mathbf{z}} = \hat{\mathbf{n}} \times \mathbf{H}$, where $\hat{\mathbf{n}} = -\hat{\mathbf{z}}$ is the outward normal to the conductor.

This result provides a justification of the boundary condition $\mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H}$ at an interface with a perfect conductor. □

5.4 Single Dielectric Slab

Multiple interface problems can be handled in a straightforward way with the help of the matching and propagation matrices. For example, Fig. 5.4.1 shows a two-interface problem with a dielectric slab η_1 separating the semi-infinite media η_a and η_b.

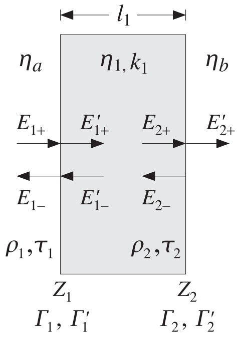


Fig. 5.4.1 Single dielectric slab.

Let l_1 be the width of the slab, $k_1 = \omega/c_1$ the propagation wavenumber, and $\lambda_1 = 2\pi/k_1$ the corresponding wavelength within the slab. We have $\lambda_1 = \lambda_0/n_1$, where λ_0 is the free-space wavelength and n_1 the refractive index of the slab. We assume the incident field is from the left medium η_a, and thus, in medium η_b there is only a forward wave.

Let ρ_1, ρ_2 be the elementary reflection coefficients from the left sides of the two interfaces, and let τ_1, τ_2 be the corresponding transmission coefficients:

$$\rho_1 = \frac{\eta_1 - \eta_a}{\eta_1 + \eta_a}\,, \quad \rho_2 = \frac{\eta_b - \eta_1}{\eta_b + \eta_1}\,, \quad \tau_1 = 1 + \rho_1\,, \quad \tau_2 = 1 + \rho_2 \tag{5.4.1}$$

To determine the reflection coefficient Γ_1 into medium η_a, we apply Eq. (5.2.9) to relate Γ_1 to the reflection coefficient Γ_1' at the right-side of the first interface. Then, we propagate to the left of the second interface with Eq. (5.1.12) to get:

$$\Gamma_1 = \frac{\rho_1 + \Gamma_1'}{1 + \rho_1 \Gamma_1'} = \frac{\rho_1 + \Gamma_2 e^{-2jk_1 l_1}}{1 + \rho_1 \Gamma_2 e^{-2jk_1 l_1}} \tag{5.4.2}$$

At the second interface, we apply Eq. (5.2.9) again to relate Γ_2 to Γ_2'. Because there are no backward-moving waves in medium η_b, we have $\Gamma_2' = 0$. Thus,

$$\Gamma_2 = \frac{\rho_2 + \Gamma_2'}{1 + \rho_2 \Gamma_2'} = \rho_2$$

We finally find for Γ_1:

$$\boxed{\Gamma_1 = \frac{\rho_1 + \rho_2 e^{-2jk_1 l_1}}{1 + \rho_1 \rho_2 e^{-2jk_1 l_1}}} \tag{5.4.3}$$

This expression can be thought of as function of frequency. Assuming a lossless medium η_1, we have $2k_1 l_1 = \omega(2l_1/c_1) = \omega T$, where $T = 2l_1/c_1 = 2(n_1 l_1)/c_0$ is the *two-way travel time delay* through medium η_1. Thus, we can write:

$$\boxed{\Gamma_1(\omega) = \frac{\rho_1 + \rho_2 e^{-j\omega T}}{1 + \rho_1 \rho_2 e^{-j\omega T}}} \tag{5.4.4}$$

This can also be expressed as a z-transform. Denoting the two-way travel time delay in the z-domain by $z^{-1} = e^{-j\omega T} = e^{-2jk_1 l_1}$, we may rewrite Eq. (5.4.4) as the first-order digital filter *transfer function*:

$$\boxed{\Gamma_1(z) = \frac{\rho_1 + \rho_2 z^{-1}}{1 + \rho_1 \rho_2 z^{-1}}} \tag{5.4.5}$$

An alternative way to derive Eq. (5.4.3) is working with wave impedances, which are continuous across interfaces. The wave impedance at interface-2 is $Z_2 = Z_2'$, but $Z_2' = \eta_b$ because there is no backward wave in medium η_b. Thus, $Z_2 = \eta_b$. Using the propagation equation for impedances, we find:

$$Z_1 = Z_1' = \eta_1 \frac{Z_2 + j\eta_1 \tan k_1 l_1}{\eta_1 + jZ_2 \tan k_1 l_1} = \eta_1 \frac{\eta_b + j\eta_1 \tan k_1 l_1}{\eta_1 + j\eta_b \tan k_1 l_1}$$

Inserting this into $\Gamma_1 = (Z_1 - \eta_a)/(Z_1 + \eta_a)$ gives Eq. (5.4.3). Working with wave impedances is always more convenient if the interfaces are positioned at half- or quarter-wavelength spacings.

If we wish to determine the overall *transmission response* into medium η_b, that is, the quantity $\mathcal{T} = E_{2+}'/E_{1+}$, then we must work with the matrix formulation. Starting at the left interface and successively applying the matching and propagation matrices, we obtain:

$$\begin{aligned}
\begin{bmatrix} E_{1+} \\ E_{1-} \end{bmatrix} &= \frac{1}{\tau_1}\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{bmatrix} E_{1+}' \\ E_{1-}' \end{bmatrix} = \frac{1}{\tau_1}\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{bmatrix} e^{jk_1 l_1} & 0 \\ 0 & e^{-jk_1 l_1} \end{bmatrix}\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix} \\
&= \frac{1}{\tau_1}\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{bmatrix} e^{jk_1 l_1} & 0 \\ 0 & e^{-jk_1 l_1} \end{bmatrix}\frac{1}{\tau_2}\begin{bmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{bmatrix}\begin{bmatrix} E_{2+}' \\ 0 \end{bmatrix}
\end{aligned}$$

where we set $E_{2-}' = 0$ by assumption. Multiplying the matrix factors out, we obtain:

$$\begin{aligned}
E_{1+} &= \frac{e^{jk_1 l_1}}{\tau_1 \tau_2}\big(1 + \rho_1 \rho_2 e^{-2jk_1 l_1}\big)E_{2+}' \\
E_{1-} &= \frac{e^{jk_1 l_1}}{\tau_1 \tau_2}\big(\rho_1 + \rho_2 e^{-2jk_1 l_1}\big)E_{2+}'
\end{aligned}$$

These may be solved for the reflection and transmission responses:

$$\begin{aligned}
\Gamma_1 &= \frac{E_{1-}}{E_{1+}} = \frac{\rho_1 + \rho_2 e^{-2jk_1 l_1}}{1 + \rho_1 \rho_2 e^{-2jk_1 l_1}} \\
\mathcal{T} &= \frac{E_{2+}'}{E_{1+}} = \frac{\tau_1 \tau_2 e^{-jk_1 l_1}}{1 + \rho_1 \rho_2 e^{-2jk_1 l_1}}
\end{aligned} \tag{5.4.6}$$

The transmission response has an overall delay factor of $e^{-jk_1 l_1} = e^{-j\omega T/2}$, representing the *one-way* travel time delay through medium η_1.

For convenience, we summarize the match-and-propagate equations relating the field quantities at the *left* of interface-1 to those at the *left* of interface-2. The forward and backward electric fields are related by the *transfer matrix*:

$$\begin{aligned}
\begin{bmatrix} E_{1+} \\ E_{1-} \end{bmatrix} &= \frac{1}{\tau_1}\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{bmatrix} e^{jk_1 l_1} & 0 \\ 0 & e^{-jk_1 l_1} \end{bmatrix}\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix} \\
\begin{bmatrix} E_{1+} \\ E_{1-} \end{bmatrix} &= \frac{1}{\tau_1}\begin{bmatrix} e^{jk_1 l_1} & \rho_1 e^{-jk_1 l_1} \\ \rho_1 e^{jk_1 l_1} & e^{-jk_1 l_1} \end{bmatrix}\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix}
\end{aligned} \tag{5.4.7}$$

The reflection responses are related by Eq. (5.4.2):

$$\Gamma_1 = \frac{\rho_1 + \Gamma_2 e^{-2jk_1 l_1}}{1 + \rho_1 \Gamma_2 e^{-2jk_1 l_1}} \tag{5.4.8}$$

The total electric and magnetic fields at the two interfaces are continuous across the interfaces and are related by Eq. (5.1.13):

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} \cos k_1 l_1 & j\eta_1 \sin k_1 l_1 \\ j\eta_1^{-1} \sin k_1 l_1 & \cos k_1 l_1 \end{bmatrix}\begin{bmatrix} E_2 \\ H_2 \end{bmatrix} \tag{5.4.9}$$

Eqs. (5.4.7)–(5.4.9) are valid in general, regardless of what is to the right of the second interface. There could be a semi-infinite uniform medium or any combination of multiple slabs. These equations were simplified in the single-slab case because we assumed that there was a uniform medium to the right and that there were no backward-moving waves.

For lossless media, energy conservation states that the energy flux into medium η_1 must equal the energy flux out of it. It is equivalent to the following relationship between Γ and $\mathcal{T}$, which can proved using Eq. (5.4.6):

$$\frac{1}{\eta_a}\left(1-|\Gamma_1|^2\right)=\frac{1}{\eta_b}\,|\mathcal{T}|^2 \tag{5.4.10}$$

Thus, if we call $|\Gamma_1|^2$ the *reflectance* of the slab, representing the fraction of the incident power that gets reflected back into medium η_a, then the quantity

$$\boxed{1-|\Gamma_1|^2=\frac{\eta_a}{\eta_b}\,|\mathcal{T}|^2=\frac{n_b}{n_a}|\mathcal{T}|^2} \tag{5.4.11}$$

will be the *transmittance* of the slab, representing the fraction of the incident power that gets transmitted through into the right medium η_b. The presence of the factors η_a, η_b can be can be understood as follows:

$$\frac{\mathcal{P}_{\text{transmitted}}}{\mathcal{P}_{\text{incident}}}=\frac{\dfrac{1}{2\eta_b}|E'_{2+}|^2}{\dfrac{1}{2\eta_a}|E_{1+}|^2}=\frac{\eta_a}{\eta_b}\,|\mathcal{T}|^2$$

5.5 *Reflectionless Slab*

The *zeros* of the transfer function (5.4.5) correspond to a reflectionless interface. Such zeros can be realized exactly only in two special cases, that is, for slabs that have either half-wavelength or quarter-wavelength thickness. It is evident from Eq. (5.4.5) that a zero will occur if $\rho_1+\rho_2 z^{-1}=0$, which gives the condition:

$$z=e^{2jk_1l_1}=-\frac{\rho_2}{\rho_1} \tag{5.5.1}$$

Because the right-hand side is real-valued and the left-hand side has unit magnitude, this condition can be satisfied only in the following two cases:

$$\begin{aligned}
&z=e^{2jk_1l_1}=1, &&\rho_2=-\rho_1, &&\text{(half-wavelength thickness)}\\
&z=e^{2jk_1l_1}=-1, &&\rho_2=\rho_1, &&\text{(quarter-wavelength thickness)}
\end{aligned}$$

The first case requires that $2k_1l_1$ be an integral multiple of 2π, that is, $2k_1l_1=2m\pi$, where m is an integer. This gives the *half-wavelength* condition $l_1=m\lambda_1/2$, where λ_1 is the wavelength in medium-1. In addition, the condition $\rho_2=-\rho_1$ requires that:

$$\frac{\eta_b-\eta_1}{\eta_b+\eta_1}=\rho_2=-\rho_1=\frac{\eta_a-\eta_1}{\eta_a+\eta_1}\quad\Leftrightarrow\quad\eta_a=\eta_b$$

that is, the media to the left and right of the slab must be the *same*. The second possibility requires $e^{2jk_1l_1}=-1$, or that $2k_1l_1$ be an odd multiple of π, that is, $2k_1l_1=(2m+1)\pi$, which translates into the *quarter-wavelength* condition $l_1=(2m+1)\lambda_1/4$. Furthermore, the condition $\rho_2=\rho_1$ requires:

$$\frac{\eta_b-\eta_1}{\eta_b+\eta_1}=\rho_2=\rho_1=\frac{\eta_1-\eta_a}{\eta_1+\eta_a}\quad\Leftrightarrow\quad\eta_1^2=\eta_a\eta_b$$

To summarize, a reflectionless slab, $\Gamma_1=0$, can be realized only in the two cases:

$$\begin{aligned}
&\text{half-wave:} &&l_1=m\,\frac{\lambda_1}{2}, &&\eta_1\ \text{arbitrary}, &&\eta_a=\eta_b\\
&\text{quarter-wave:} &&l_1=(2m+1)\,\frac{\lambda_1}{4}, &&\eta_1=\sqrt{\eta_a\eta_b}, &&\eta_a,\eta_b\ \text{arbitrary}
\end{aligned} \tag{5.5.2}$$

An equivalent way of stating these conditions is to say that the *optical length* of the slab must be a half or quarter of the *free-space* wavelength λ_0. Indeed, if n_1 is the refractive index of the slab, then its optical length is n_1l_1, and in the half-wavelength case we have $n_1l_1=n_1m\lambda_1/2=m\lambda_0/2$, where we used $\lambda_1=\lambda_0/n_1$. Similarly, we have $n_1l_1=(2m+1)\lambda_0/4$ in the quarter-wavelength case. In terms of the refractive indices, Eq. (5.5.2) reads:

$$\boxed{\begin{aligned}
&\text{half-wave:} &&n_1l_1=m\,\frac{\lambda_0}{2}, &&n_1\ \text{arbitrary}, &&n_a=n_b\\
&\text{quarter-wave:} &&n_1l_1=(2m+1)\,\frac{\lambda_0}{4}, &&n_1=\sqrt{n_an_b}, &&n_a,n_b\ \text{arbitrary}
\end{aligned}} \tag{5.5.3}$$

The reflectionless matching condition can also be derived by working with wave impedances. For half-wavelength spacing, we have from Eq. (5.1.18) $Z_1=Z_2=\eta_b$. The condition $\Gamma_1=0$ requires $Z_1=\eta_a$, thus, matching occurs if $\eta_a=\eta_b$. Similarly, for the quarter-wavelength case, we have $Z_1=\eta_1^2/Z_2=\eta_1^2/\eta_b=\eta_a$.

We emphasize that the reflectionless response $\Gamma_1=0$ is obtained only at certain slab widths (half- or quarter-wavelength), or equivalently, at certain operating frequencies. These operating frequencies correspond to $\omega T=2m\pi$, or, $\omega T=(2m+1)\pi$, that is, $\omega=2m\pi/T=m\omega_0$, or, $\omega=(2m+1)\omega_0/2$, where we defined $\omega_0=2\pi/T$.

The dependence on l_1 or ω can be seen from Eq. (5.4.5). For the half-wavelength case, we substitute $\rho_2=-\rho_1$ and for the quarter-wavelength case, $\rho_2=\rho_1$. Then, the reflection transfer functions become:

$$\begin{aligned}
\Gamma_1(z)&=\frac{\rho_1(1-z^{-1})}{1-\rho_1^2z^{-1}}, \quad \text{(half-wave)}\\
\Gamma_1(z)&=\frac{\rho_1(1+z^{-1})}{1+\rho_1^2z^{-1}}, \quad \text{(quarter-wave)}
\end{aligned} \tag{5.5.4}$$

where $z=e^{2jk_1l_1}=e^{j\omega T}$. The magnitude-square responses then take the form:

$$\begin{aligned}
|\Gamma_1|^2&=\frac{2\rho_1^2\big(1-\cos(2k_1l_1)\big)}{1-2\rho_1^2\cos(2k_1l_1)+\rho_1^4}=\frac{2\rho_1^2(1-\cos\omega T)}{1-2\rho_1^2\cos\omega T+\rho_1^4}, \quad \text{(half-wave)}\\
|\Gamma_1|^2&=\frac{2\rho_1^2\big(1+\cos(2k_1l_1)\big)}{1+2\rho_1^2\cos(2k_1l_1)+\rho_1^4}=\frac{2\rho_1^2(1+\cos\omega T)}{1+2\rho_1^2\cos\omega T+\rho_1^4}, \quad \text{(quarter-wave)}
\end{aligned} \tag{5.5.5}$$

These expressions are periodic in l_1 with period $\lambda_1/2$, and periodic in ω with period $\omega_0=2\pi/T$. In DSP language, the slab acts as a digital filter with sampling frequency ω_0. The maximum reflectivity occurs at $z=-1$ and $z=1$ for the half- and quarter-wavelength cases. The maximum squared responses are in either case:

$$|\Gamma_1|^2_{\max} = \frac{4\rho_1^2}{(1+\rho_1^2)^2}$$

Fig. 5.5.1 shows the magnitude responses for the three values of the reflection coefficient: $|\rho_1| = 0.9$, 0.7, and 0.5. The closer ρ_1 is to unity, the narrower are the reflectionless notches.

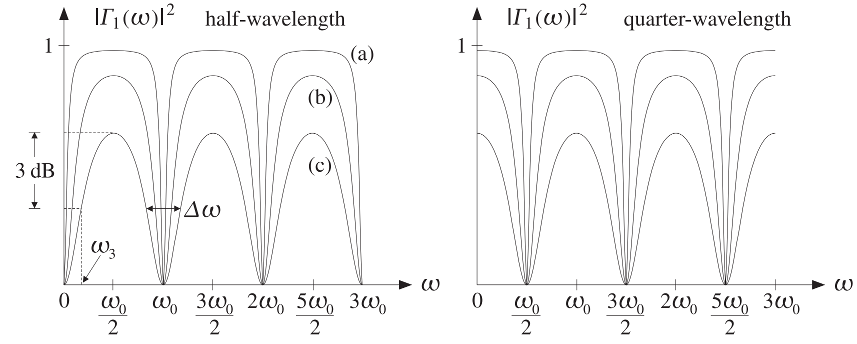


Fig. 5.5.1 Reflection responses $|\Gamma(\omega)|^2$. (a) $|\rho_1| = 0.9$, (b) $|\rho_1| = 0.7$, (c) $|\rho_1| = 0.5$.

It is evident from these figures that for the same value of ρ_1, the half- and quarter-wavelength cases have the same notch widths. A standard measure for the width is the 3-dB width, which for the half-wavelength case is twice the 3-dB frequency ω_3, that is, $\Delta\omega = 2\omega_3$, as shown in Fig. 5.5.1 for the case $|\rho_1| = 0.5$. The frequency ω_3 is determined by the 3-dB half-power condition:

$$|\Gamma_1(\omega_3)|^2 = \frac{1}{2}\,|\Gamma_1|^2_{\max}$$

or, equivalently:

$$\frac{2\rho_1^2(1-\cos\omega_3 T)}{1-2\rho_1^2\cos\omega_3 T+\rho_1^4} = \frac{1}{2}\,\frac{4\rho_1^2}{(1+\rho_1^2)^2}$$

Solving for the quantity $\cos\omega_3 T = \cos(\Delta\omega T/2)$, we find:

$$\cos\left(\frac{\Delta\omega T}{2}\right) = \frac{2\rho_1^2}{1+\rho_1^4} \quad\Leftrightarrow\quad \tan\left(\frac{\Delta\omega T}{4}\right) = \frac{1-\rho_1^2}{1+\rho_1^2} \tag{5.5.6}$$

If ρ_1^2 is very near unity, then $1-\rho_1^2$ and $\Delta\omega$ become small, and we may use the approximation $\tan x \simeq x$ to get:

$$\frac{\Delta\omega T}{4} \simeq \frac{1-\rho_1^2}{1+\rho_1^2} \simeq \frac{1-\rho_1^2}{2}$$

which gives the approximation:

$$\Delta\omega T = 2(1-\rho_1^2) \tag{5.5.7}$$

This is a standard approximation for digital filters relating the 3-dB width of a pole peak to the radius of the pole [48]. For any desired value of the bandwidth $\Delta\omega$, Eq. (5.5.6) or (5.5.7) may be thought of as a design condition that determines ρ_1.

Fig. 5.5.2 shows the corresponding *transmittances* $1-|\Gamma_1(\omega)|^2$ of the slabs. The transmission response acts as a periodic bandpass filter. This is the simplest example of a so-called *Fabry-Perot interference filter* or *Fabry-Perot resonator*. Such filters find application in the spectroscopic analysis of materials. We discuss them further in Chap. 6.

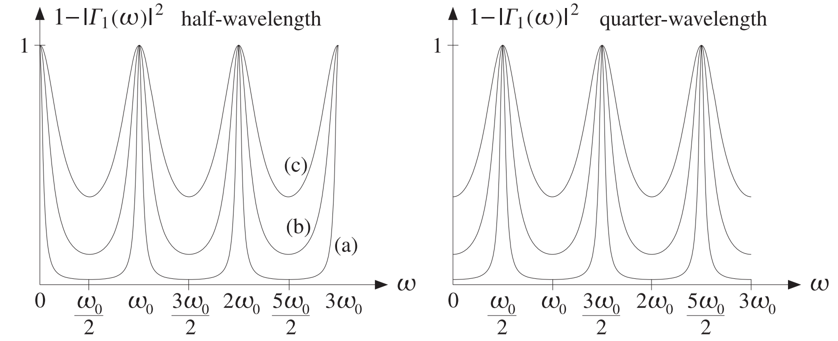


Fig. 5.5.2 Transmittance of half- and quarter-wavelength dielectric slab.

Using Eq. (5.5.5), we may express the frequency response of the half-wavelength transmittance filter in the following equivalent forms:

$$1-|\Gamma_1(\omega)|^2 = \frac{(1-\rho_1^2)^2}{1-2\rho_1^2\cos\omega T+\rho_1^4} = \frac{1}{1+\mathcal{F}\sin^2(\omega T/2)} \tag{5.5.8}$$

where the $\mathcal{F}$ is called the *finesse* in the Fabry-Perot context and is defined by:

$$\mathcal{F} = \frac{4\rho_1^2}{(1-\rho_1^2)^2}$$

The finesse is a measure of the peak width, with larger values of $\mathcal{F}$ corresponding to narrower peaks. The connection of $\mathcal{F}$ to the 3-dB width (5.5.6) is easily found to be:

$$\tan\left(\frac{\Delta\omega T}{4}\right) = \frac{1-\rho_1^2}{1+\rho_1^2} = \frac{1}{\sqrt{1+\mathcal{F}}} \tag{5.5.9}$$

Quarter-wavelength slabs may be used to design anti-reflection coatings for lenses, so that all incident light on a lens gets through. Half-wavelength slabs, which require that the medium be the same on either side of the slab, may be used in designing radar domes (radomes) protecting microwave antennas, so that the radiated signal from the antenna goes through the radome wall without getting reflected back towards the antenna.

Example 5.5.1: Determine the reflection coefficients of half- and quarter-wave slabs that do not necessarily satisfy the impedance conditions of Eq. (5.5.2).

Solution: The reflection response is given in general by Eq. (5.4.6). For the half-wavelength case, we have $e^{2jk_1l_1} = 1$ and we obtain:

$$\Gamma_1 = \frac{\rho_1 + \rho_2}{1 + \rho_1\rho_2} = \frac{\dfrac{\eta_1 - \eta_a}{\eta_1 + \eta_a} + \dfrac{\eta_b - \eta_1}{\eta_b + \eta_1}}{1 + \dfrac{\eta_1 - \eta_a}{\eta_1 + \eta_a}\,\dfrac{\eta_b - \eta_1}{\eta_b + \eta_1}} = \frac{\eta_b - \eta_a}{\eta_b + \eta_a} = \frac{n_a - n_b}{n_a + n_b}$$

This is the same as if the slab were absent. For this reason, half-wavelength slabs are sometimes referred to as *absentee* layers. Similarly, in the quarter-wavelength case, we have $e^{2jk_1l_1} = -1$ and find:

$$\Gamma_1 = \frac{\rho_1 - \rho_2}{1 - \rho_1\rho_2} = \frac{\eta_1^2 - \eta_a\eta_b}{\eta_1^2 + \eta_a\eta_b} = \frac{n_an_b - n_1^2}{n_an_b + n_1^2}$$

The slab becomes reflectionless if the conditions (5.5.2) are satisfied. □

Example 5.5.2: *Antireflection Coating.* Determine the refractive index of a quarter-wave antireflection coating on a glass substrate with index 1.5.

Solution: From Eq. (5.5.3), we have with $n_a = 1$ and $n_b = 1.5$:

$$n_1 = \sqrt{n_an_b} = \sqrt{1.5} = 1.22$$

The closest refractive index that can be obtained is that of cryolite ($\mathrm{Na_3AlF_6}$) with $n_1 = 1.35$ and magnesium fluoride ($\mathrm{MgF_2}$) with $n_1 = 1.38$. Magnesium fluoride is usually preferred because of its durability. Such a slab will have a reflection coefficient as given by the previous example:

$$\Gamma_1 = \frac{\rho_1 - \rho_2}{1 - \rho_1\rho_2} = \frac{\eta_1^2 - \eta_a\eta_b}{\eta_1^2 + \eta_a\eta_b} = \frac{n_an_b - n_1^2}{n_an_b + n_1^2} = \frac{1.5 - 1.38^2}{1.5 + 1.38^2} = -0.118$$

with reflectance $|\Gamma|^2 = 0.014$, or 1.4 percent. This is to be compared to the 4 percent reflectance of uncoated glass that we determined in Example 5.3.1.

Fig. 5.5.3 shows the reflectance $|\Gamma(\lambda)|^2$ as a function of the free-space wavelength λ. The reflectance remains less than one or two percent in the two cases, over almost the entire visible spectrum.

The slabs were designed to have quarter-wavelength thickness at $\lambda_0 = 550$ nm, that is, the optical length was $n_1l_1 = \lambda_0/4$, resulting in $l_1 = 112.71$ nm and 99.64 nm in the two cases of $n_1 = 1.22$ and $n_1 = 1.38$. Such extremely thin dielectric films are fabricated by means of a thermal evaporation process [592,594].

The MATLAB code used to generate this example was as follows:

```
n = [1, 1.22, 1.50]; L = 1/4;                refractive indices and optical length
lambda = linspace(400,700,101) / 550;        visible spectrum wavelengths
Gamma1 = multidiel(n, L, lambda);            reflection response of slab
```

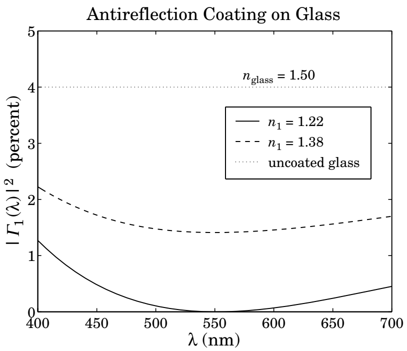


Fig. 5.5.3 Reflectance over the visible spectrum.

The syntax and use of the function `multidiel` is discussed in Sec. 6.1. The dependence of Γ on λ comes through the quantity $k_1l_1 = 2\pi(n_1l_1)/\lambda$. Since $n_1l_1 = \lambda_0/4$, we have $k_1l_1 = 0.5\pi\lambda_0/\lambda$. □

Example 5.5.3: *Thick Glasses.* Interference phenomena, such as those arising from the multiple reflections within a slab, are not observed if the slabs are “thick” (compared to the wavelength.) For example, typical glass windows seem perfectly transparent.

If one had a glass plate of thickness, say, of $l = 1.5$ mm and index $n = 1.5$, it would have optical length $nl = 1.5\times1.5 = 2.25$ mm $= 225\times10^4$ nm. At an operating wavelength of $\lambda_0 = 450$ nm, the glass plate would act as a half-wave transparent slab with $nl = 10^4(\lambda_0/2)$, that is, 10^4 half-wavelengths long.

Such plate would be very difficult to construct as it would require that l be built with an accuracy of a few percent of $\lambda_0/2$. For example, assuming $n(\Delta l) = 0.01(\lambda_0/2)$, the plate should be constructed with an accuracy of one part in a million: $\Delta l/l = n\Delta l/(nl) = 0.01/10^4 = 10^{-6}$. (That is why thin films are constructed by a carefully controlled evaporation process.)

More realistically, a typical glass plate can be constructed with an accuracy of one part in a thousand, $\Delta l/l = 10^{-3}$, which would mean that within the manufacturing uncertainty Δl, there would still be ten half-wavelengths, $n\Delta\lambda = 10^{-3}(nl) = 10(\lambda_0/2)$.

The overall power reflection response will be obtained by *averaging* $|\Gamma_1|^2$ over several $\lambda_0/2$ cycles, such as the above ten. Because of periodicity, the average of $|\Gamma_1|^2$ over several cycles is the same as the average over one cycle, that is,

$$\overline{|\Gamma_1|^2} = \frac{1}{\omega_0}\int_0^{\omega_0} |\Gamma_1(\omega)|^2\, d\omega$$

where $\omega_0 = 2\pi/T$ and T is the two-way travel-time delay. Using either of the two expressions in Eq. (5.5.5), this integral can be done exactly resulting in the average reflectance and transmittance:

$$\overline{|\Gamma_1|^2} = \frac{2\rho_1^2}{1+\rho_1^2}\,, \quad 1 - \overline{|\Gamma_1|^2} = \frac{1-\rho_1^2}{1+\rho_1^2} = \frac{2n}{n^2+1} \tag{5.5.10}$$

where we used $\rho_1 = (1-n)/(1+n)$. This explains why glass windows do not exhibit a frequency-selective behavior as predicted by Eq. (5.5.5). For $n = 1.5$, we find $1 - \overline{|\Gamma_1|^2} = 0.9231$, that is, 92.31% of the incident light is transmitted through the plate.

The same expressions for the average reflectance and transmittance can be obtained by summing incoherently all the multiple reflections within the slab, that is, summing the multiple reflections of power instead of field amplitudes. The timing diagram for such multiple reflections is shown in Fig. 5.6.1.

Indeed, if we denote by $p_r = \rho_1^2$ and $p_t = 1 - p_r = 1 - \rho_1^2$, the power reflection and transmission coefficients, then the first reflection of power will be p_r. The power transmitted through the left interface will be p_t and through the second interface p_t^2 (assuming the same medium to the right.) The reflected power at the second interface will be $p_t p_r$ and will come back and transmit through the left interface giving $p_t^2 p_r$.

Similarly, after a second round trip, the reflected power will be $p_t^2 p_r^3$, while the transmitted power to the right of the second interface will be $p_t^2 p_r^2$, and so on. Summing up all the reflected powers to the left and those transmitted to the right, we find:

$$\overline{|\Gamma_1|^2} = p_r + p_t^2 p_r + p_t^2 p_r^3 + p_t^2 p_r^5 + \cdots = p_r + \frac{p_t^2 p_r}{1 - p_r^2} = \frac{2p_r}{1 + p_r}$$

$$1 - \overline{|\Gamma_1|^2} = p_t^2 + p_t^2 p_r^2 + p_t^2 p_r^4 + \cdots = \frac{p_t^2}{1 - p_r^2} = \frac{1 - p_r}{1 + p_r}$$

where we used $p_t = 1 - p_r$. These are equivalent to Eqs. (5.5.10). □

Example 5.5.4: *Radomes.* A radome protecting a microwave transmitter has $\epsilon = 4\epsilon_0$ and is designed as a half-wavelength reflectionless slab at the operating frequency of 10 GHz. Determine its thickness.

Next, suppose that the operating frequency is 1% off its nominal value of 10 GHz. Calculate the percentage of reflected power back towards the transmitting antenna.

Determine the operating *bandwidth* as that frequency interval about the 10 GHz operating frequency within which the reflected power remains at least 30 dB below the incident power.

Solution: The free-space wavelength is $\lambda_0 = c_0/f_0 = 30$ GHz cm/10 GHz = 3 cm. The refractive index of the slab is $n = 2$ and the wavelength inside it, $\lambda_1 = \lambda_0/n = 3/2 = 1.5$ cm. Thus, the slab thickness will be the half-wavelength $l_1 = \lambda_1/2 = 0.75$ cm, or any other integral multiple of this.

Assume now that the operating frequency is $\omega = \omega_0 + \delta\omega$, where $\omega_0 = 2\pi f_0 = 2\pi/T$. Denoting $\delta = \delta\omega/\omega_0$, we can write $\omega = \omega_0(1+\delta)$. The numerical value of δ is very small, $\delta = 1\% = 0.01$. Therefore, we can do a first-order calculation in δ. The reflection coefficient ρ_1 and reflection response Γ are:

$$\rho_1 = \frac{\eta - \eta_0}{\eta + \eta_0} = \frac{0.5 - 1}{0.5 + 1} = -\frac{1}{3}\,, \qquad \Gamma_1(\omega) = \frac{\rho_1(1 - z^{-1})}{1 - \rho_1^2 z^{-1}} = \frac{\rho_1(1 - e^{-j\omega T})}{1 - \rho_1^2 e^{-j\omega T}}$$

where we used $\eta = \eta_0/n = \eta_0/2$. Noting that $\omega T = \omega_0 T(1+\delta) = 2\pi(1+\delta)$, we can expand the delay exponential to first-order in δ:

$$z^{-1} = e^{-j\omega T} = e^{-2\pi j(1+\delta)} = e^{-2\pi j} e^{-2\pi j\delta} = e^{-2\pi j\delta} \simeq 1 - 2\pi j\delta$$

Thus, the reflection response becomes to first-order in δ:

$$\Gamma_1 \simeq \frac{\rho_1\big(1 - (1 - 2\pi j\delta)\big)}{1 - \rho_1^2(1 - 2\pi j\delta)} = \frac{\rho_1 2\pi j\delta}{1 - \rho_1^2 + \rho_1^2 2\pi j\delta} \simeq \frac{\rho_1 2\pi j\delta}{1 - \rho_1^2}$$

where we replaced the denominator by its zeroth-order approximation because the numerator is already first-order in δ. It follows that the power reflection response will be:

$$|\Gamma_1|^2 = \frac{\rho_1^2(2\pi\delta)^2}{(1 - \rho_1^2)^2}$$

Evaluating this expression for $\delta = 0.01$ and $\rho_1 = -1/3$, we find $|\Gamma|^2 = 0.00049$, or 0.049 percent of the incident power gets reflected. Next, we find the frequency about ω_0 at which the reflected power is $A = 30$ dB below the incident power. Writing again, $\omega = \omega_0 + \delta\omega = \omega_0(1+\delta)$ and assuming δ is small, we have the condition:

$$|\Gamma_1|^2 = \frac{\rho_1^2(2\pi\delta)^2}{(1 - \rho_1^2)^2} = \frac{\mathcal{P}_{\text{refl}}}{\mathcal{P}_{\text{inc}}} = 10^{-A/10} \quad \Rightarrow \quad \delta = \frac{1 - \rho_1^2}{2\pi|\rho_1|}\, 10^{-A/20}$$

Evaluating this expression, we find $\delta = 0.0134$, or $\delta\omega = 0.0134\omega_0$. The bandwidth will be twice that, $\Delta\omega = 2\delta\omega = 0.0268\omega_0$, or in Hz, $\Delta f = 0.0268 f_0 = 268$ MHz. □

Example 5.5.5: Because of manufacturing imperfections, suppose that the actual constructed thickness of the above radome is 1% off the desired half-wavelength thickness. Determine the percentage of reflected power in this case.

Solution: This is essentially the same as the previous example. Indeed, the quantity $\theta = \omega T = 2k_1 l_1 = 2\omega l_1/c_1$ can change either because of ω or because of l_1. A simultaneous infinitesimal change (about the nominal value $\theta_0 = \omega_0 T = 2\pi$) will give:

$$\delta\theta = 2(\delta\omega)l_1/c_1 + 2\omega_0(\delta l_1)/c_1 \quad \Rightarrow \quad \delta = \frac{\delta\theta}{\theta_0} = \frac{\delta\omega}{\omega_0} + \frac{\delta l_1}{l_1}$$

In the previous example, we varied ω while keeping l_1 constant. Here, we vary l_1, while keeping ω constant, so that $\delta = \delta l_1/l_1$. Thus, we have $\delta\theta = \theta_0\delta = 2\pi\delta$. The corresponding delay factor becomes approximately $z^{-1} = e^{-j\theta} = e^{-j(2\pi + \delta\theta)} = 1 - j\delta\theta = 1 - 2\pi j\delta$. The resulting expression for the power reflection response is identical to the above and its numerical value is the same if $\delta = 0.01$. □

Example 5.5.6: Because of weather conditions, suppose that the characteristic impedance of the medium outside the above radome is 1% off the impedance inside. Calculate the percentage of reflected power in this case.

Solution: Suppose that the outside impedance changes to $\eta_b = \eta_0 + \delta\eta$. The wave impedance at the outer interface will be $Z_2 = \eta_b = \eta_0 + \delta\eta$. Because the slab length is still a half-wavelength, the wave impedance at the inner interface will be $Z_1 = Z_2 = \eta_0 + \delta\eta$. It follows that the reflection response will be:

$$\Gamma_1 = \frac{Z_1 - \eta_0}{Z_1 + \eta_0} = \frac{\eta_0 + \delta\eta - \eta_0}{\eta_0 + \delta\eta + \eta_0} = \frac{\delta\eta}{2\eta_0 + \delta\eta} \simeq \frac{\delta\eta}{2\eta_0}$$

where we replaced the denominator by its zeroth-order approximation in $\delta\eta$. Evaluating at $\delta\eta/\eta_0 = 1\% = 0.01$, we find $\Gamma_1 = 0.005$, which leads to a reflected power of $|\Gamma_1|^2 = 2.5\times 10^{-5}$, or, 0.0025 percent. □

5.6 Time-Domain Reflection Response

We conclude our discussion of the single slab by trying to understand its behavior in the time domain. The z-domain reflection transfer function of Eq. (5.4.5) incorporates the effect of all multiple reflections that are set up within the slab as the wave bounces back and forth at the left and right interfaces. Expanding Eq. (5.4.5) in a partial fraction expansion and then in power series in z^{-1} gives:

$$\Gamma_1(z)=\frac{\rho_1+\rho_2 z^{-1}}{1+\rho_1\rho_2 z^{-1}}=\frac{1}{\rho_1}-\frac{1}{\rho_1}\,\frac{(1-\rho_1^2)}{1+\rho_1\rho_2 z^{-1}}=\rho_1+\sum_{n=1}^{\infty}(1-\rho_1^2)\,(-\rho_1)^{n-1}\rho_2^n\,z^{-n}$$

Using the reflection coefficient from the right of the first interface, $\rho_1'=-\rho_1$, and the transmission coefficients $\tau_1=1+\rho_1$ and $\tau_1'=1+\rho_1'=1-\rho_1$, we have $\tau_1\tau_1'=1-\rho_1^2$. Then, the above power series can be written as a function of frequency in the form:

$$\Gamma_1(\omega)=\rho_1+\sum_{n=1}^{\infty}\tau_1\tau_1'(\rho_1')^{n-1}\rho_2^n\,z^{-n}=\rho_1+\sum_{n=1}^{\infty}\tau_1\tau_1'(\rho_1')^{n-1}\rho_2^n\,e^{-j\omega nT}$$

where we set $z^{-1}=e^{-j\omega T}$. It follows that the time-domain *reflection impulse response*, that is, the inverse Fourier transform of $\Gamma_1(\omega)$, will be the sum of discrete impulses:

$$\Gamma_1(t)=\rho_1\delta(t)+\sum_{n=1}^{\infty}\tau_1\tau_1'(\rho_1')^{n-1}\rho_2^n\,\delta(t-nT) \tag{5.6.1}$$

This is the response of the slab to a forward-moving impulse striking the left interface at $t=0$, that is, the response to the input $E_{1+}(t)=\delta(t)$. The first term $\rho_1\delta(t)$ is the impulse immediately reflected at $t=0$ with the reflection coefficient ρ_1. The remaining terms represent the multiple reflections within the slab. Fig. 5.6.1 is a *timing diagram* that traces the reflected and transmitted impulses at the first and second interfaces.

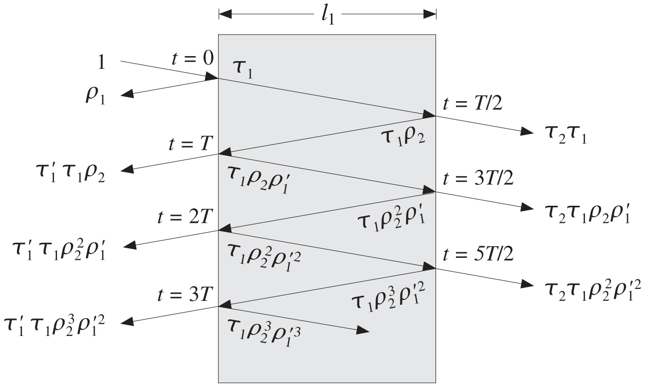

l_1
1
$t = 0$
τ_1
ρ_1
$t = T/2$
$\tau_2\tau_1$
$\tau_1\rho_2$
$t = T$
$\tau_1'\,\tau_1\rho_2$
$\tau_1\rho_2\rho_1'$
$t = 3T/2$
$\tau_2\tau_1\rho_2\rho_1'$
$\tau_1\rho_2^2\rho_1'$
$t = 2T$
$\tau_1'\,\tau_1\rho_2^2\rho_1'$
$\tau_1\rho_2^2\rho_1'^2$
$t = 5T/2$
$\tau_2\tau_1\rho_2^2\rho_1'^2$
$\tau_1\rho_2^3\rho_1'^2$
$t = 3T$
$\tau_1'\,\tau_1\rho_2^3\rho_1'^2$
$\tau_1\rho_2^3\rho_1'^3$

Fig. 5.6.1 Multiple reflections building up the reflection and transmission responses.

The input pulse $\delta(t)$ gets transmitted to the inside of the left interface and picks up a transmission coefficient factor τ_1. In $T/2$ seconds this pulse strikes the right interface and causes a reflected wave whose amplitude is changed by the reflection coefficient ρ_2 into $\tau_1\rho_2$.

Thus, the pulse $\tau_1\rho_2\delta(t-T/2)$ gets reflected backwards and will arrive at the left interface $T/2$ seconds later, that is, at time $t=T$. A proportion τ_1' of it will be transmitted through to the left, and a proportion ρ_1' will be re-reflected towards the right. Thus, at time $t=T$, the transmitted pulse into the left medium will be $\tau_1\tau_1'\rho_2\delta(t-T)$, and the re- reflected pulse $\tau_1\rho_1'\rho_2\delta(t-T)$.

The re-reflected pulse will travel forward to the right interface, arriving there at time $t=3T/2$ getting reflected backwards picking up a factor ρ_2. This will arrive at the left at time $t=2T$. The part transmitted to the left will be now $\tau_1\tau_1'\rho_1'\rho_2^2\delta(t-2T)$, and the part re-reflected to the right $\tau_1\rho_1'^2\rho_2^2\delta(t-2T)$. And so on, after the nth round trip, the pulse transmitted to the left will be $\tau_1\tau_1'(\rho_1')^{n-1}\rho_2^n\delta(t-nT)$. The sum of all the reflected pulses will be $\Gamma_1(t)$ of Eq. (5.6.1).

In a similar way, we can derive the overall *transmission response* to the right. It is seen in the figure that the transmitted pulse at time $t=nT+(T/2)$ will be $\tau_1\tau_2(\rho_1')^n\rho_2^n$. Thus, the overall transmission impulse response will be:

$$\mathcal{T}(t)=\sum_{n=0}^{\infty}\tau_1\tau_2(\rho_1')^n\rho_2^n\,\delta(t-nT-T/2)$$

It follows that its Fourier transform will be:

$$\mathcal{T}(\omega)=\sum_{n=0}^{\infty}\tau_1\tau_2(\rho_1')^n\rho_2^n e^{-jn\omega T}e^{-j\omega T/2}$$

which sums up to Eq. (5.4.6):

$$\mathcal{T}(\omega)=\frac{\tau_1\tau_2 e^{-j\omega T/2}}{1-\rho_1'\rho_2 e^{-j\omega T}}=\frac{\tau_1\tau_2 e^{-j\omega T/2}}{1+\rho_1\rho_2 e^{-j\omega T}} \tag{5.6.2}$$

For an incident field $E_{1+}(t)$ with arbitrary time dependence, the overall reflection response of the slab is obtained by convolving the impulse response $\Gamma_1(t)$ with $E_{1+}(t)$. This follows from the linear superposition of the reflection responses of all the frequency components of $E_{1+}(t)$, that is,

$$E_{1-}(t)=\int_{-\infty}^{\infty}\Gamma_1(\omega)E_{1+}(\omega)e^{j\omega t}\frac{d\omega}{2\pi}\,,\quad\text{where}\quad E_{1+}(t)=\int_{-\infty}^{\infty}E_{1+}(\omega)e^{j\omega t}\frac{d\omega}{2\pi}$$

Then, the convolution theorem of Fourier transforms implies that:

$$E_{1-}(t)=\int_{-\infty}^{\infty}\Gamma_1(\omega)E_{1+}(\omega)e^{j\omega t}\frac{d\omega}{2\pi}=\int_{-\infty}^{-\infty}\Gamma_1(t')E_{1+}(t-t')dt' \tag{5.6.3}$$

Inserting (5.6.1), we find that the reflected wave arises from the multiple reflections of $E_{1+}(t)$ as it travels and bounces back and forth between the two interfaces:

$$E_{1-}(t)=\rho_1 E_{1+}(t)+\sum_{n=1}^{\infty}\tau_1\tau_1'(\rho_1')^{n-1}\rho_2^n\,E_{1+}(t-nT) \tag{5.6.4}$$

For a causal waveform $E_{1+}(t)$, the summation over n will be finite, such that at each time $t\geq 0$ only the terms that have $t-nT\geq 0$ will be present. In a similar fashion, we find for the overall transmitted response into medium η_b:

$$E'_{2+}(t)=\int_{-\infty}^{-\infty}\mathcal{T}(t')E_{1+}(t-t')\,dt'=\sum_{n=0}^{\infty}\tau_1\tau_2(\rho'_1)^n\rho_2^n\,E_{1+}(t-nT-T/2)\tag{5.6.5}$$

We will use similar techniques later on to determine the transient responses of transmission lines.

5.7 Two Dielectric Slabs

Next, we consider more than two interfaces. As we mentioned in the previous section, Eqs. (5.4.7)–(5.4.9) are general and can be applied to all successive interfaces. Fig. 5.7.1 shows three interfaces separating four media. The overall reflection response can be calculated by successive application of Eq. (5.4.8):

$$\Gamma_1=\frac{\rho_1+\Gamma_2e^{-2jk_1l_1}}{1+\rho_1\Gamma_2e^{-2jk_1l_1}}\,,\quad \Gamma_2=\frac{\rho_2+\Gamma_3e^{-2jk_2l_2}}{1+\rho_2\Gamma_3e^{-2jk_2l_2}}$$

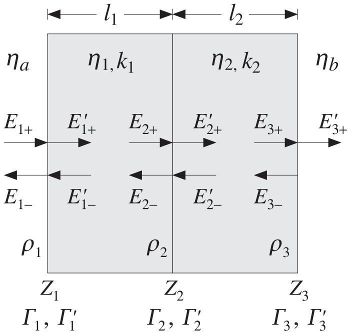


Fig. 5.7.1 Two dielectric slabs.

If there is no backward-moving wave in the right-most medium, then $\Gamma'_3=0$, which implies $\Gamma_3=\rho_3$. Substituting Γ_2 into Γ_1 and denoting $z_1=e^{2jk_1l_1}$, $z_2=e^{2jk_2l_2}$, we eventually find:

$$\Gamma_1=\frac{\rho_1+\rho_2z_1^{-1}+\rho_1\rho_2\rho_3z_2^{-1}+\rho_3z_1^{-1}z_2^{-1}}{1+\rho_1\rho_2z_1^{-1}+\rho_2\rho_3z_2^{-1}+\rho_1\rho_3z_1^{-1}z_2^{-1}}\tag{5.7.1}$$

The reflection response Γ_1 can alternatively be determined from the knowledge of the wave impedance $Z_1=E_1/H_1$ at interface-1:

$$\Gamma_1=\frac{Z_1-\eta_a}{Z_1+\eta_a}$$

The fields E_1,H_1 are obtained by successively applying Eq. (5.4.9):

$$\begin{aligned}\begin{bmatrix}E_1\\H_1\end{bmatrix}&=\begin{bmatrix}\cos k_1l_1 & j\eta_1\sin k_1l_1\\ j\eta_1^{-1}\sin k_1l_1 & \cos k_1l_1\end{bmatrix}\begin{bmatrix}E_2\\H_2\end{bmatrix}\\&=\begin{bmatrix}\cos k_1l_1 & j\eta_1\sin k_1l_1\\ j\eta_1^{-1}\sin k_1l_1 & \cos k_1l_1\end{bmatrix}\begin{bmatrix}\cos k_2l_2 & j\eta_2\sin k_2l_2\\ j\eta_2^{-1}\sin k_2l_2 & \cos k_2l_2\end{bmatrix}\begin{bmatrix}E_3\\H_3\end{bmatrix}\end{aligned}$$

But at interface-3, $E_3=E'_3=E'_{3+}$ and $H_3=Z_3^{-1}E_3=\eta_b^{-1}E'_{3+}$, because $Z_3=\eta_b$. Therefore, we can obtain the fields E_1,H_1 by the matrix multiplication:

$$\begin{bmatrix}E_1\\H_1\end{bmatrix}=\begin{bmatrix}\cos k_1l_1 & j\eta_1\sin k_1l_1\\ j\eta_1^{-1}\sin k_1l_1 & \cos k_1l_1\end{bmatrix}\begin{bmatrix}\cos k_2l_2 & j\eta_2\sin k_2l_2\\ j\eta_2^{-1}\sin k_2l_2 & \cos k_2l_2\end{bmatrix}\begin{bmatrix}1\\ \eta_b^{-1}\end{bmatrix}E'_{3+}$$

Because Z_1 is the ratio of E_1 and H_1, the factor E'_{3+} cancels out and can be set equal to unity.

Example 5.7.1: Determine Γ_1 if both slabs are quarter-wavelength slabs. Repeat if both slabs are half-wavelength and when one is half- and the other quarter-wavelength.

Solution: Because $l_1=\lambda_1/4$ and $l_2=\lambda_2/4$, we have $2k_1l_1=2k_2l_2=\pi$, and it follows that $z_1=z_2=-1$. Then, Eq. (5.7.1) becomes:

$$\Gamma_1=\frac{\rho_1-\rho_2-\rho_1\rho_2\rho_3+\rho_3}{1-\rho_1\rho_2-\rho_2\rho_3+\rho_1\rho_3}$$

A simpler approach is to work with wave impedances. Using $Z_3=\eta_b$, we have:

$$Z_1=\frac{\eta_1^2}{Z_2}=\frac{\eta_1^2}{\eta_2^2/Z_3}=\frac{\eta_1^2}{\eta_2^2}Z_3=\frac{\eta_1^2}{\eta_2^2}\eta_b$$

Inserting this into $\Gamma_1=(Z_1-\eta_a)/(Z_1+\eta_a)$, we obtain:

$$\Gamma_1=\frac{\eta_1^2\eta_b-\eta_2^2\eta_a}{\eta_1^2\eta_b+\eta_2^2\eta_a}$$

The two expressions for Γ_1 are equivalent. The input impedance Z_1 can also be obtained by matrix multiplication. Because $k_1l_1=k_2l_2=\pi/2$, we have $\cos k_1l_1=0$ and $\sin k_1l_1=1$ and the propagation matrices for E_1,H_1 take the simplified form:

$$\begin{bmatrix}E_1\\H_1\end{bmatrix}=\begin{bmatrix}0 & j\eta_1\\ j\eta_1^{-1} & 0\end{bmatrix}\begin{bmatrix}0 & j\eta_2\\ j\eta_2^{-1} & 0\end{bmatrix}\begin{bmatrix}1\\ \eta_b^{-1}\end{bmatrix}E'_{3+}=\begin{bmatrix}-\eta_1\eta_2^{-1}\\ -\eta_2\eta_1^{-1}\eta_b^{-1}\end{bmatrix}E'_{3+}$$

The ratio E_1/H_1 gives the same answer for Z_1 as above. When both slabs are half-wavelength, the impedances propagate unchanged: $Z_1=Z_2=Z_3$, but $Z_3=\eta_b$.

If η_1 is half- and η_2 quarter-wavelength, then, $Z_1=Z_2=\eta_2^2/Z_3=\eta_2^2/\eta_b$. And, if the quarter-wavelength is first and the half-wavelength second, $Z_1=\eta_1^2/Z_2=\eta_1^2/Z_3=\eta_1^2/\eta_b$. The corresponding reflection coefficient Γ_1 is in the three cases:

$$\Gamma_1=\frac{\eta_b-\eta_a}{\eta_b+\eta_a}\,,\quad \Gamma_1=\frac{\eta_2^2-\eta_a\eta_b}{\eta_2^2+\eta_a\eta_b}\,,\quad \Gamma_1=\frac{\eta_1^2-\eta_a\eta_b}{\eta_1^2+\eta_a\eta_b}$$

These expressions can also be derived by Eq. (5.7.1), or by the matrix method. □

The frequency dependence of Eq. (5.7.1) arises through the factors z_1, z_2, which can be written in the forms: $z_1 = e^{j\omega T_1}$ and $z_2 = e^{j\omega T_2}$, where $T_1 = 2l_1/c_1$ and $T_2 = 2l_2/c_2$ are the two-way travel time delays through the two slabs.

A case of particular interest arises when the slabs are designed to have the *equal* travel-time delays so that $T_1 = T_2 \equiv T$. Then, defining a common variable $z = z_1 = z_2 = e^{j\omega T}$, we can write the reflection response as a second-order digital filter transfer function:

$$\Gamma_1(z) = \frac{\rho_1 + \rho_2(1+\rho_1\rho_3)z^{-1} + \rho_3 z^{-2}}{1 + \rho_2(\rho_1+\rho_3)z^{-1} + \rho_1\rho_3 z^{-2}} \tag{5.7.2}$$

In the next chapter, we discuss further the properties of such higher-order reflection transfer functions arising from multilayer dielectric slabs.

5.8 Reflection by a Moving Boundary

Reflection and transmission by moving boundaries, such as reflection from a moving mirror, introduce Doppler shifts in the frequencies of the reflected and transmitted waves. Here, we look at the problem of normal incidence on a dielectric interface that is moving with constant velocity v perpendicularly to the interface, that is, along the z-direction as shown in Fig. 5.8.1. Additional examples may be found in [435–453]. The case of oblique incidence is discussed in Sec. 7.12.

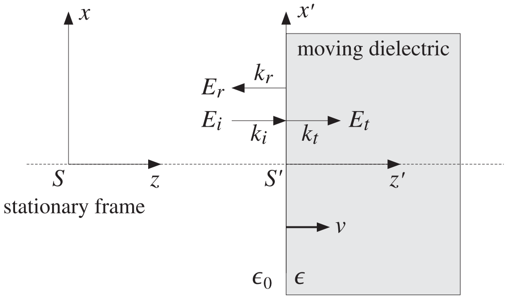


Fig. 5.8.1 Reflection and transmission at a moving boundary.

The dielectric is assumed to be non-magnetic and lossless with permittivity ϵ. The left medium is free space ϵ_0. The electric field is assumed to be in the x-direction and thus, the magnetic field will be in the y-direction. We consider two coordinate frames, the fixed frame S with coordinates $\{t, x, y, z\}$, and the moving frame S' with $\{t', x', y', z'\}$. The two sets of coordinates are related by the Lorentz transformation equations (H.1) of Appendix H.

We are interested in determining the Doppler-shifted frequencies of the reflected and transmitted waves, as well as the reflection and transmission coefficients as measured in the fixed frame S.

The procedure for solving this type of problem—originally suggested by Einstein in his 1905 special relativity paper [435]—is to solve the reflection and transmission problem in the moving frame S' with respect to which the boundary is at rest, and then transform the results back to the fixed frame S using the Lorentz transformation properties of the fields. In the fixed frame S, the fields to the left and right of the interface will have the forms:

$$\text{left}\;\begin{cases} E_x = E_i e^{j(\omega t - k_i z)} + E_r e^{j(\omega_r t + k_r z)} \\ H_y = H_i e^{j(\omega t - k_i z)} - H_r e^{j(\omega_r t + k_r z)} \end{cases} \qquad \text{right}\;\begin{cases} E_x = E_t e^{j(\omega_t t - k_t z)} \\ H_y = H_t e^{j(\omega_t t - k_t z)} \end{cases} \tag{5.8.1}$$

where $\omega, \omega_r, \omega_t$ and k_i, k_r, k_t are the frequencies and wavenumbers of the incident, reflected, and transmitted waves measured in S. Because of Lorentz invariance, the propagation phases remain unchanged in the frames S and S', that is,

$$\begin{aligned} \phi_i &= \omega t - k_i z = \omega' t' - k_i' z' = \phi_i' \\ \phi_r &= \omega_r t + k_r z = \omega' t' + k_r' z' = \phi_r' \\ \phi_t &= \omega_t t - k_t z = \omega' t' - k_t' z' = \phi_t' \end{aligned} \tag{5.8.2}$$

In the frame S' where the dielectric is at rest, all three frequencies are the same and set equal to ω'. This is a consequence of the usual tangential boundary conditions applied to the interface at rest. Note that ϕ_r can be written as $\phi_r = \omega_r t - (-k_r)z$ implying that the reflected wave is propagating in the negative z-direction. In the rest frame S' of the boundary, the wavenumbers are:

$$k_i' = \frac{\omega'}{c}\,, \quad k_r' = \frac{\omega'}{c}\,, \quad k_t' = \omega'\sqrt{\epsilon\mu_0} = n\frac{\omega'}{c} \tag{5.8.3}$$

where c is the speed of light in vacuum and $n = \sqrt{\epsilon/\epsilon_0}$ is the refractive index of the dielectric at rest. The frequencies and wavenumbers in the fixed frame S are related to those in S' by applying the Lorentz transformation of Eq. (H.14) to the frequency-wavenumber four-vectors $(\omega/c, 0, 0, k_i)$, $(\omega_r/c, 0, 0, -k_r)$, and $(\omega_t/c, 0, 0, k_t)$:

$$\begin{aligned} \omega &= \gamma(\omega' + \beta c k_i') = \omega'\gamma(1+\beta) \\ k_i &= \gamma\Big(k_i' + \frac{\beta}{c}\omega'\Big) = \frac{\omega'}{c}\gamma(1+\beta) \\ \omega_r &= \gamma\big(\omega' + \beta c(-k_r')\big) = \omega'\gamma(1-\beta) \\ -k_r &= \gamma\Big(-k_r' + \frac{\beta}{c}\omega'\Big) = -\frac{\omega'}{c}\gamma(1-\beta) \\ \omega_t &= \gamma(\omega' + \beta c k_t') = \omega'\gamma(1+\beta n) \\ k_t &= \gamma\Big(k_t' + \frac{\beta}{c}\omega'\Big) = \frac{\omega'}{c}\gamma(n+\beta) \end{aligned} \tag{5.8.4}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. Eliminating the primed quantities, we obtain the Doppler-shifted frequencies of the reflected and transmitted waves:

$$\boxed{\omega_r = \omega\frac{1-\beta}{1+\beta}}\,, \qquad \boxed{\omega_t = \omega\frac{1+\beta n}{1+\beta}} \tag{5.8.5}$$

The phase velocities of the incident, reflected, and transmitted waves are:

$$v_i = \frac{\omega}{k_i} = c\,, \quad v_r = \frac{\omega_r}{k_r} = c\,, \quad v_t = \frac{\omega_t}{k_t} = c\frac{1+\beta n}{n+\beta} \tag{5.8.6}$$

These can also be derived by applying Einstein's velocity addition theorem of Eq. (H.8). For example, we have for the transmitted wave:

$$v_t = \frac{v_d + v}{1 + v_d v/c^2} = \frac{c/n + v}{1 + (c/n)v/c^2} = c\frac{1+\beta n}{n+\beta}$$

where $v_d = c/n$ is the phase velocity within the dielectric at rest. To first-order in $\beta = v/c$, the phase velocity within the moving dielectric becomes:

$$v_t = c\frac{1+\beta n}{n+\beta} \simeq \frac{c}{n} + v\left(1 - \frac{1}{n^2}\right)$$

The second term is known as the "Fresnel drag." The quantity $n_t = (n+\beta)/(1+\beta n)$ may be thought of as the "effective" refractive index of the moving dielectric as measured in the fixed system S.

Next, we derive the reflection and transmission coefficients. In the rest-frame S' of the dielectric, the fields have the usual forms derived earlier in Sections 5.1 and 5.2:

$$\text{left}\;\begin{cases} E'_x = E'_i\left(e^{j\phi'_i} + \rho e^{j\phi'_r}\right) \\[4pt] H'_y = \dfrac{1}{\eta_0}E'_i\left(e^{j\phi'_i} - \rho e^{j\phi'_r}\right)\end{cases} \qquad \text{right}\;\begin{cases} E'_x = \tau E'_i e^{j\phi'_t} \\[4pt] H'_y = \dfrac{1}{\eta}\tau E'_i e^{j\phi'_t}\end{cases} \tag{5.8.7}$$

where

$$\eta = \frac{\eta_0}{n}\,, \quad \rho = \frac{\eta - \eta_0}{\eta + \eta_0} = \frac{1-n}{1+n}\,, \quad \tau = 1 + \rho = \frac{2}{1+n}$$

The primed fields can be transformed to the fixed frame S using the inverse of the Lorentz transformation equations (H.31), that is,

$$\begin{aligned} E_x &= \gamma(E'_x + \beta c B'_y) = \gamma(E'_x + \beta\eta_0 H'_y) \\ H_y &= \gamma(H'_y + c\beta D'_x) = \gamma(H'_y + c\beta\epsilon E'_x) \end{aligned} \tag{5.8.8}$$

where we replaced $B'_y = \mu_0 H'_y$, $c\mu_0 = \eta_0$, and $D'_x = \epsilon E'_x$ (of course, $\epsilon = \epsilon_0$ in the left medium). Using the invariance of the propagation phases, we find for the fields at the left side of the interface:

$$E_x = \gamma\big[E'_i(e^{j\phi_i} + \rho e^{j\phi_r}) + \beta E'_i(e^{j\phi_i} - \rho e^{j\phi_r})\big] = E'_i\gamma\big[(1+\beta)e^{j\phi_i} + \rho(1-\beta)e^{j\phi_r}\big] \tag{5.8.9}$$

Similarly, for the right side of the interface we use the property $\eta_0/\eta = n$ to get:

$$E_x = \gamma\big[\tau E'_i e^{j\phi_t} + \beta n\tau E'_i e^{j\phi_t}\big] = \gamma\tau E'_i(1+\beta n)e^{j\phi_t} \tag{5.8.10}$$

Comparing these with Eq. (5.8.1), we find the incident, reflected, and transmitted electric field amplitudes:

$$E_i = \gamma E'_i(1+\beta)\,, \quad E_r = \rho\gamma E'_i(1-\beta)\,, \quad E_t = \tau\gamma E'_i(1+\beta n) \tag{5.8.11}$$

from which we obtain the reflection and transmission coefficients in the fixed frame S:

$$\boxed{\frac{E_r}{E_i} = \rho\frac{1-\beta}{1+\beta}}\,, \qquad \boxed{\frac{E_t}{E_i} = \tau\frac{1+\beta n}{1+\beta}} \tag{5.8.12}$$

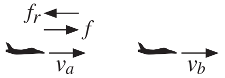
The case of a perfect mirror is also covered by these expressions by setting $\rho = -1$ and $\tau = 0$. Eq. (5.8.5) is widely used in Doppler radar applications. Typically, the boundary (the target) is moving at non-relativistic speeds so that $\beta = v/c \ll 1$. In such case, the first-order approximation of (5.8.5) is adequate:

$$f_r \simeq f(1 - 2\beta) = f\left(1 - 2\frac{v}{c}\right) \quad\Rightarrow\quad \frac{\Delta f}{f} = -2\frac{v}{c} \tag{5.8.13}$$

where $\Delta f = f_r - f$ is the Doppler shift. The negative sign means that $f_r < f$ if the target is receding away from the source of the wave, and $f_r > f$ if it is approaching the source.

As we mentioned in Sec. 2.11, if the source of the wave is moving with velocity v_a and the target with velocity v_b (with respect to a common fixed frame, such as the ground), then one must use the relative velocity $v = v_b - v_a$ in the above expression:

$$\boxed{\frac{\Delta f}{f} = \frac{f_r - f}{f} = 2\frac{v_a - v_b}{c}} \tag{5.8.14}$$



5.9 Problems

5.1 Fill in the details of the equivalence between Eq. (5.2.2) and (5.2.3), that is,

$$\begin{aligned} E_+ + E_- &= E'_+ + E'_- \\ \frac{1}{\eta}(E_+ - E_-) &= \frac{1}{\eta'}(E'_+ - E'_-) \end{aligned} \quad\Leftrightarrow\quad \begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \frac{1}{\tau}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} E'_+ \\ E'_- \end{bmatrix}$$

5.2 Fill in the details of the equivalences stated in Eq. (5.2.9), that is,

$$Z = Z' \quad\Leftrightarrow\quad \Gamma = \frac{\rho + \Gamma'}{1 + \rho\Gamma'} \quad\Leftrightarrow\quad \Gamma' = \frac{\rho' + \Gamma}{1 + \rho'\Gamma}$$

Show that if there is no left-incident field from the right, then $\Gamma = \rho$, and if there is no right-incident field from the left, then, $\Gamma' = 1/\rho'$. Explain the asymmetry of the two cases.

5.3 Let ρ, τ be the reflection and transmission coefficients from the left side of an interface and let ρ', τ' be those from the right, as defined in Eq. (5.2.5). One of the two media may be lossy, and therefore, its characteristic impedance and hence ρ, τ may be complex-valued. Show and interpret the relationships:

$$1 - |\rho|^2 = \text{Re}\left(\frac{\eta}{\eta'}\right)|\tau|^2 = \text{Re}(\tau^*\tau')$$

5.4 Show that the reflection and transmission responses of the single dielectric slab of Fig. 5.4.1 are given by Eq. (5.4.6), that is,

$$\Gamma = \frac{\rho_1 + \rho_2 e^{-2jk_1 l_1}}{1 + \rho_1\rho_2 e^{-2jk_1 l_1}}\,, \qquad \mathcal{T} = \frac{E'_{2+}}{E_{1+}} = \frac{\tau_1\tau_2 e^{-jk_1 l_1}}{1 + \rho_1\rho_2 e^{-2jk_1 l_1}}$$

Moreover, using these expressions show and interpret the relationship:

$$\frac{1}{\eta_a}\big(1-|\Gamma|^2\big)=\frac{1}{\eta_b}\,|\mathcal{T}|^2$$

5.5 A 1-GHz plane wave is incident normally onto a thick copper plate ($\sigma = 5.8\times 10^7$ S/m.) Can the plate be considered to be a good conductor at this frequency? Calculate the percentage of the incident power that enters the plate. Calculate the attenuation coefficient within the conductor and express it in units of dB/m. What is the penetration depth in mm?

5.6 With the help of Fig. 5.5.1, argue that the 3-dB width $\Delta\omega$ is related to the 3-dB frequency ω_3 by $\Delta\omega = 2\omega_3$ and $\Delta\omega = \omega_0 - 2\omega_3$, in the cases of half- and quarter-wavelength slabs. Then, show that ω_3 and $\Delta\omega$ are given by:

$$\cos\omega_3 T = \pm\frac{2\rho_1^2}{1+\rho_1^4}\,,\qquad \tan\left(\frac{\Delta\omega T}{4}\right)=\frac{1-\rho_1^2}{1+\rho_1^2}$$

5.7 A fiberglass ($\epsilon = 4\epsilon_0$) radome protecting a microwave antenna is designed as a half-wavelength reflectionless slab at the operating frequency of 12 GHz.

a. Determine three possible thicknesses (in cm) for this radome.

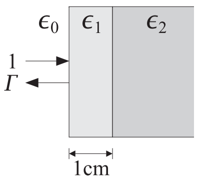
b. Determine the 15-dB and 30-dB bandwidths in GHz about the 12 GHz operating frequency , defined as the widths over which the reflected power is 15 or 30 dB below the incident power.

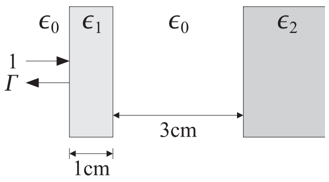
5.8 A 5 GHz wave is normally incident from air onto a dielectric slab of thickness of 1 cm and refractive index of 1.5, as shown below. The medium to the right of the slab has an index of 2.25.

a. Write an analytical expression of the reflectance $|\Gamma(f)|^2$ as a function of frequency and sketch it versus f over the interval $0 \le f \le 15$ GHz. What is the value of the reflectance at 5 GHz?

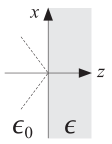
b. Next, the 1-cm slab is moved to the left by a distance of 3 cm, creating an air-gap between it and the rightmost dielectric. Repeat all the questions of part (a).

c. Repeat part (a), if the slab thickness is 2 cm.





5.9 A single-frequency plane wave is incident obliquely from air onto a planar interface with a medium of permittivity $\epsilon = 2\epsilon_0$, as shown below. The incident wave has the following phasor form:



$$\mathbf{E}(z)=\left(\frac{\hat{\mathbf{x}}+\hat{\mathbf{z}}}{\sqrt{2}}+j\,\hat{\mathbf{y}}\right)e^{-jk(z-x)/\sqrt{2}} \tag{5.9.1}$$

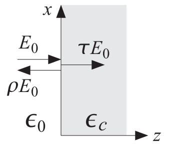
a. Determine the angle of incidence θ in degrees and decide which of the two dashed lines in the figure represents the incident wave. Moreover, determine the angle of refraction θ' in degrees and indicate the refracted wave's direction on the figure below.

b. Write an expression for the reflected wave that is similar to Eq. (5.9.1), but also includes the dependence on the TE and TM Fresnel reflection coefficients (please evaluate these coefficients numerically.) Similarly, give an expression for the transmitted wave.

c. Determine the polarization type (circular, elliptic, left, right, linear, etc.) of the incident wave and of the reflected wave.

5.10 A uniform plane wave is incident normally on a planar interface, as shown below. The medium to the left of the interface is air, and the medium to the right is lossy with an effective complex permittivity ϵ_c, complex wavenumber $k' = \beta' - j\alpha' = \omega\sqrt{\mu_0\epsilon_c}$, and complex characteristic impedance $\eta_c = \sqrt{\mu_0/\epsilon_c}$. The electric field to the left and right of the interface has the following form:

$$E_x=\begin{cases}E_0e^{-jkz}+\rho E_0e^{jkz}, & z\le 0\\ \tau E_0 e^{-jk'z}, & z\ge 0\end{cases}$$



where ρ, τ are the reflection and transmission coefficients.

1. Determine the magnetic field at both sides of the interface.
2. Show that the Poynting vector only has a z-component, given as follows at the two sides of the interface:

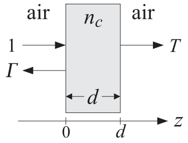
$$\mathcal{P}=\frac{|E_0|^2}{2\eta_0}\big(1-|\rho|^2\big)\,,\qquad \mathcal{P}'=\frac{|E_0|^2}{2\omega\mu_0}\beta'|\tau|^2e^{-2\alpha' z}$$

3. Moreover, show that $\mathcal{P}=\mathcal{P}'$ at the interface, (i.e., at $z = 0$).

5.11 Consider a lossy dielectric slab of thickness d and complex refractive index $n_c = n_r - jn_i$ at an operating frequency ω, with air on both sides as shown below.

a. Let $k = \beta - j\alpha = k_0 n_c$ and $\eta_c = \eta_0/n_c$ be the corresponding complex wavenumber and characteristic impedance of the slab, where $k_0 = \omega\sqrt{\mu_0\epsilon_0} = \omega/c_0$ and $\eta_0 = \sqrt{\mu_0/\epsilon_0}$. Show that the transmission response of the slab may be expressed as follows:

$$T=\frac{1}{\cos kd + j\,\dfrac{1}{2}\left(n_c+\dfrac{1}{n_c}\right)\sin kd}$$



b. At the cell phone frequency of 900 MHz, the complex refractive index of concrete is $n_c = 2.5 - 0.14j$. Calculate the percentage of the transmitted power through a 20-cm concrete wall. How is this percentage related to T and why?

c. Is there anything interesting about the choice $d = 20$ cm? Explain.

5.12 Consider the slab of the previous problem. The tangential electric field has the following form in the three regions $z \le 0$, $0 \le z \le d$, and $z \ge d$:

$$E(z)=\begin{cases} e^{-jk_0 z}+\Gamma e^{jk_0 z}\,, & \text{if} \quad z \le 0 \\ Ae^{-jkz}+Be^{jkz}\,, & \text{if} \quad 0 \le z \le d \\ Te^{-jk_0(z-d)}\,, & \text{if} \quad z \ge d \end{cases}$$

where k_0 and k were defined in the previous problem.

a. What are the corresponding expressions for the magnetic field $H(z)$?

b. Set up and solve four equations from which the four unknowns Γ, A, B, T may be determined.

c. If the slab is lossless and is designed to be a half-wave slab at the frequency ω, then what is the value of T?

d. If the slab is is lossy with $n_c = n_r - jn_i$ and is designed to be a half-wave slab with respect to the real part β of k, that is, $\beta d = \pi$, then, show that T is given by:

$$T = -\frac{1}{\cosh\alpha d + \dfrac{1}{2}\left(n_c + \dfrac{1}{n_c}\right)\sinh\alpha d}$$

5.13 Consider a two-layer dielectric structure as shown in Fig. 5.7.1, and let n_a, n_1, n_2, n_b be the refractive indices of the four media. Consider the four cases: (a) both layers are quarter-wave, (b) both layers are half-wave, (c) layer-1 is quarter- and layer-2 half-wave, and (d) layer-1 is half- and layer-2 quarter-wave. Show that the reflection coefficient at interface-1 is given by the following expressions in the four cases:

$$\Gamma_1 = \frac{n_a n_2^2 - n_b n_1^2}{n_a n_2^2 + n_b n_1^2}\,, \quad \Gamma_1 = \frac{n_a - n_b}{n_a + n_b}\,, \quad \Gamma_1 = \frac{n_a n_b - n_1^2}{n_a n_b + n_1^2}\,, \quad \Gamma_1 = \frac{n_a n_b - n_2^2}{n_a n_b + n_2^2}$$

5.14 Consider the lossless two-slab structure of Fig. 5.7.1. Write down all the transfer matrices relating the fields $E_{i\pm}$, $i = 1,2,3$ at the left sides of the three interfaces. Then, show the energy conservation equations:

$$\frac{1}{\eta_a}\left(|E_{1+}|^2 - |E_{1-}|^2\right) = \frac{1}{\eta_1}\left(|E_{2+}|^2 - |E_{2-}|^2\right) = \frac{1}{\eta_2}\left(|E_{3+}|^2 - |E_{3-}|^2\right) = \frac{1}{\eta_b}\,|E'_{3+}|^2$$

5.15 An alternative way of representing the propagation relationship Eq. (5.1.12) is in terms of the hyperbolic w-plane variable defined in terms of the reflection coefficient Γ, or equivalently, the wave impedance Z as follows:

$$\Gamma = e^{-2w} \quad \Leftrightarrow \quad Z = \eta\coth(w) \tag{5.9.2}$$

Show the equivalence of these expressions. Writing $\Gamma_1 = e^{-2w_1}$ and $\Gamma_2 = e^{-2w_2}$, show that Eq. (5.1.12) becomes equivalent to:

$$w_1 = w_2 + jkl \qquad \text{(propagation in } w\text{-domain)} \tag{5.9.3}$$

This form is essentially the mathematical (as opposed to graphical) version of the *Smith chart* and is particularly useful for numerical computations using MATLAB.

5.16 Plane A flying at a speed of 900 km/hr with respect to the ground is approaching plane B. Plane A's Doppler radar, operating at the X-band frequency of 10 GHz, detects a positive Doppler shift of 2 kHz in the return frequency. Determine the speed of plane B with respect to the ground. [*Ans.* 792 km/hr.]

5.17 The complete set of Lorentz transformations of the fields in Eq. (5.8.8) is as follows (see also Eq. (H.31) of Appendix H):

$$E_x = \gamma(E'_x + \beta c B'_y)\,, \quad H_y = \gamma(H'_y + c\beta D'_x)\,, \quad D_x = \gamma(D'_x + \frac{1}{c}\beta H'_y)\,, \quad B_y = \gamma(B'_y + \frac{1}{c}\beta E'_x)$$

The constitutive relations in the rest frame S' of the moving dielectric are the usual ones, that is, $B'_y = \mu H'_y$ and $D'_x = \epsilon E'_x$. By eliminating the primed quantities in terms of the unprimed ones, show that the constitutive relations have the following form in the fixed system S:

$$D_x = \frac{(1-\beta^2)\epsilon E_x - \beta(n^2-1)H_y/c}{1-\beta^2 n^2}\,, \quad B_y = \frac{(1-\beta^2)\mu H_y - \beta(n^2-1)E_x/c}{1-\beta^2 n^2}$$

where n is the refractive index of the moving medium, $n = \sqrt{\epsilon\mu/\epsilon_0\mu_0}$. Show that for free space, the constitutive relations remain the same as in the frame S'.

6

Multilayer Structures

Higher-order transfer functions of the type of Eq. (5.7.2) can achieve broader reflectionless notches and are used in the design of thin-film antireflection coatings, dielectric mirrors, and optical interference filters [592–654,713–746], and in the design of broadband terminations of transmission lines [781–791].

They are also used in the analysis, synthesis, and simulation of fiber Bragg gratings [747–767], in the design of narrow-band transmission filters for wavelength-division multiplexing (WDM), and in other fiber-optic signal processing systems [777–780].

They are used routinely in making acoustic tube models for the analysis and synthesis of speech, with the layer recursions being mathematically equivalent to the Levinson lattice recursions of linear prediction [792–798]. The layer recursions are also used in speech recognition, disguised as the Schur algorithm.

They also find application in geophysical deconvolution and inverse scattering problems for oil exploration [799–808].

The layer recursions—known as the Schur recursions in this context—are intimately connected to the mathematical theory of lossless bounded real functions in the z-plane and positive real functions in the s-plane and find application in network analysis, synthesis, and stability [812–826].

6.1 Multiple Dielectric Slabs

The general case of arbitrary number of dielectric slabs of arbitrary thicknesses is shown in Fig. 6.1.1. There are M slabs, $M+1$ interfaces, and $M+2$ dielectric media, including the left and right semi-infinite media η_a and η_b.

The incident and reflected fields are considered at the *left* of each interface. The overall reflection response, $\Gamma_1 = E_{1-}/E_{1+}$, can be obtained recursively in a variety of ways, such as by the propagation matrices, the propagation of the impedances at the interfaces, or the propagation of the reflection responses.

The elementary reflection coefficients ρ_i from the left of each interface are defined in terms of the characteristic impedances or refractive indices as follows:

$$\rho_i = \frac{\eta_i - \eta_{i-1}}{\eta_i + \eta_{i-1}} = \frac{n_{i-1} - n_i}{n_{i-1} + n_i}\,, \quad i = 1, 2, \dots, M+1 \tag{6.1.1}$$

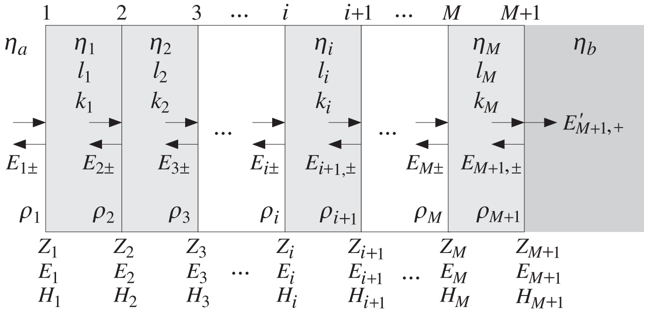


Fig. 6.1.1 Multilayer dielectric slab structure.

where $\eta_i = \eta_0/n_i$, and we must use the convention $n_0 = n_a$ and $n_{M+1} = n_b$, so that $\rho_1 = (n_a - n_1)/(n_a + n_1)$ and $\rho_{M+1} = (n_M - n_b)/(n_M + n_b)$. The forward/backward fields at the left of interface i are related to those at the left of interface $i+1$ by:

$$\begin{bmatrix} E_{i+} \\ E_{i-} \end{bmatrix} = \frac{1}{\tau_i}\begin{bmatrix} e^{jk_il_i} & \rho_i e^{-jk_il_i} \\ \rho_i e^{jk_il_i} & e^{-jk_il_i} \end{bmatrix}\begin{bmatrix} E_{i+1,+} \\ E_{i+1,-} \end{bmatrix}, \quad i = M, M-1, \dots, 1 \qquad (6.1.2)$$

where $\tau_i = 1 + \rho_i$ and k_il_i is the phase thickness of the ith slab, which can be expressed in terms of its optical thickness n_il_i and the operating free-space wavelength by $k_il_i = 2\pi(n_il_i)/\lambda$. Assuming no backward waves in the right-most medium, these recursions are initialized at the $(M+1)$st interface as follows:

$$\begin{bmatrix} E_{M+1,+} \\ E_{M+1,-} \end{bmatrix} = \frac{1}{\tau_{M+1}}\begin{bmatrix} 1 & \rho_{M+1} \\ \rho_{M+1} & 1 \end{bmatrix}\begin{bmatrix} E'_{M+1,+} \\ 0 \end{bmatrix} = \frac{1}{\tau_{M+1}}\begin{bmatrix} 1 \\ \rho_{M+1} \end{bmatrix} E'_{M+1,+}$$

It follows that the reflection responses $\Gamma_i = E_{i-}/E_{i+}$ will satisfy the recursions:

$$\boxed{\Gamma_i = \frac{\rho_i + \Gamma_{i+1}e^{-2jk_il_i}}{1 + \rho_i\Gamma_{i+1}e^{-2jk_il_i}}}, \qquad i = M, M-1, \dots, 1 \qquad (6.1.3)$$

and initialized by $\Gamma_{M+1} = \rho_{M+1}$. Similarly the recursions for the total electric and magnetic fields, which are continuous across each interface, are given by:

$$\begin{bmatrix} E_i \\ H_i \end{bmatrix} = \begin{bmatrix} \cos k_il_i & j\eta_i \sin k_il_i \\ j\eta_i^{-1}\sin k_il_i & \cos k_il_i \end{bmatrix}\begin{bmatrix} E_{i+1} \\ H_{i+1} \end{bmatrix}, \quad i = M, M-1, \dots, 1 \qquad (6.1.4)$$

and initialized at the $(M+1)$st interface as follows:

$$\begin{bmatrix} E_{M+1} \\ H_{M+1} \end{bmatrix} = \begin{bmatrix} 1 \\ \eta_b^{-1} \end{bmatrix} E'_{M+1,+}$$

It follows that the impedances at the interfaces, $Z_i = E_i/H_i$, satisfy the recursions:

$$\boxed{Z_i = \eta_i\frac{Z_{i+1} + j\eta_i\tan k_il_i}{\eta_i + jZ_{i+1}\tan k_il_i}}, \quad i = M, M-1, \dots, 1 \qquad (6.1.5)$$

and initialized by $Z_{M+1} = \eta_b$. The objective of all these recursions is to obtain the overall reflection response Γ_1 into medium η_a.

The MATLAB function `multidiel` implements the recursions (6.1.3) for such a *multi-dielectric* structure and evaluates Γ_1 and Z_1 at any desired set of free-space wavelengths. Its usage is as follows:

```
[Gamma1,Z1] = multidiel(n,L,lambda);            % multilayer dielectric structure
```

where n, L are the vectors of refractive indices of the $M+2$ media and the optical thicknesses of the M slabs, that is, in the notation of Fig. 6.1.1:

$$n = [n_a, n_1, n_2, \dots, n_M, n_b], \quad L = [n_1l_1, n_2l_2, \dots, n_Ml_M]$$

and λ is a vector of free-space wavelengths at which to evaluate Γ_1. Both the optical lengths L and the wavelengths λ are in units of some desired reference wavelength, say λ_0, typically chosen at the center of the desired band. The usage of `multidiel` was illustrated in Example 5.5.2. Additional examples are given in the next sections.

The layer recursions (6.1.2)–(6.1.5) remain essentially unchanged in the case of oblique incidence (with appropriate redefinitions of the impedances η_i) and are discussed in Chap. 7.

Next, we apply the layer recursions to the analysis and design of antireflection coatings and dielectric mirrors.

6.2 Antireflection Coatings

The simplest example of antireflection coating is the quarter-wavelength layer discussed in Example 5.5.2. Its primary drawback is that it requires the layer's refractive index to satisfy the reflectionless condition $n_1 = \sqrt{n_an_b}$.

For a typical glass substrate with index $n_b = 1.50$, we have $n_1 = 1.22$. Materials with n_1 near this value, such as magnesium fluoride with $n_1 = 1.38$, will result into some, but minimized, reflection compared to the uncoated glass case, as we saw in Example 5.5.2.

The use of multiple layers can improve the reflectionless properties of the single quarter-wavelength layer, while allowing the use of real materials. In this section, we consider three such examples.

Assuming a magnesium fluoride film and adding between it and the glass another film of higher refractive index, it is possible to achieve a reflectionless structure (at a single wavelength) by properly adjusting the film thicknesses [594,619].

With reference to the notation of Fig. 5.7.1, we have $n_a = 1$, $n_1 = 1.38$, n_2 to be determined, and $n_b = n_{\text{glass}} = 1.5$. The reflection response at interface-1 is related to the response at interface-2 by the layer recursions:

$$\Gamma_1 = \frac{\rho_1 + \Gamma_2e^{-2jk_1l_1}}{1 + \rho_1\Gamma_2e^{-2jk_1l_1}}, \quad \Gamma_2 = \frac{\rho_2 + \rho_3e^{-2jk_2l_2}}{1 + \rho_2\rho_3e^{-2jk_2l_2}}$$

The reflectionless condition is $\Gamma_1 = 0$ at an operating free-space wavelength λ_0. This requires that $\rho_1 + \Gamma_2 e^{-2jk_1l_1} = 0$, which can be written as:

$$e^{2jk_1l_1} = -\frac{\Gamma_2}{\rho_1} \tag{6.2.1}$$

Because the left-hand side has unit magnitude, we must have the condition $|\Gamma_2| = |\rho_1|$, or, $|\Gamma_2|^2 = \rho_1^2$, which is written as:

$$\left|\frac{\rho_2 + \rho_3 e^{-2jk_2l_2}}{1 + \rho_2\rho_3 e^{-2jk_2l_2}}\right|^2 = \frac{\rho_2^2 + \rho_3^2 + 2\rho_2\rho_3\cos 2k_2l_2}{1 + \rho_2^2\rho_3^2 + 2\rho_2\rho_3\cos 2k_2l_2} = \rho_1^2$$

This can be solved for $\cos 2k_2l_2$:

$$\cos 2k_2l_2 = \frac{\rho_1^2(1 + \rho_2^2\rho_3^2) - (\rho_2^2 + \rho_3^2)}{2\rho_2\rho_3(1 - \rho_1^2)}$$

Using the identity, $\cos 2k_2l_2 = 2\cos^2 k_2l_2 - 1$, we also find:

$$\begin{aligned}\cos^2 k_2l_2 &= \frac{\rho_1^2(1 - \rho_2\rho_3)^2 - (\rho_2 - \rho_3)^2}{4\rho_2\rho_3(1 - \rho_1^2)}\\ \sin^2 k_2l_2 &= \frac{(\rho_2 + \rho_3)^2 - \rho_1^2(1 + \rho_2\rho_3)^2}{4\rho_2\rho_3(1 - \rho_1^2)}\end{aligned} \tag{6.2.2}$$

It is evident from these expressions that not every combination of ρ_1, ρ_2, ρ_3 will admit a solution because the left-hand sides are positive and less than one. If we assume that $n_2 > n_1$ and $n_2 > n_b$, then, we will have $\rho_2 < 0$ and $\rho_3 > 0$. Then, it is necessary that the numerators of above expressions be negative, resulting into the conditions:

$$\left|\frac{\rho_3 + \rho_2}{1 + \rho_2\rho_3}\right|^2 < \rho_1^2 < \left|\frac{\rho_3 - \rho_2}{1 - \rho_2\rho_3}\right|^2$$

The left inequality requires that $\sqrt{n_b} < n_1 < n_b$, which is satisfied with the choices $n_1 = 1.38$ and $n_b = 1.5$. Similarly, the right inequality is violated—and therefore there is no solution—if $\sqrt{n_b} < n_2 < n_1\sqrt{n_b}$, which has the numerical range $1.22 < n_2 < 1.69$.

Catalan [594,619] used bismuth oxide (Bi_2O_3) with $n_2 = 2.45$, which satisfies the above conditions for the existence of solution. With this choice, the reflection coefficients are $\rho_1 = -0.16$, $\rho_2 = -0.28$, and $\rho_3 = 0.24$. Solving Eq. (6.2.2) for k_2l_2 and then Eq. (6.2.1) for k_1l_1, we find:

$$k_1l_1 = 2.0696, \quad k_2l_2 = 0.2848 \quad \text{(radians)}$$

Writing $k_1l_1 = 2\pi(n_1l_1)/\lambda_0$, we find the optical lengths:

$$n_1l_1 = 0.3294\lambda_0, \quad n_2l_2 = 0.0453\lambda_0$$

Fig. 6.2.1 shows the resulting reflection response Γ_1 as a function of the free-space wavelength λ, with λ_0 chosen to correspond to the middle of the visible spectrum, $\lambda_0 = 550$ nm. The figure also shows the responses of the single quarter-wave slab of Example 5.5.2.

The reflection responses were computed with the help of the MATLAB function `multidiel`. The MATLAB code used to implement this example was as follows:

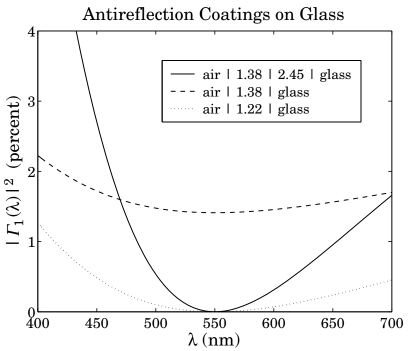


Fig. 6.2.1 Two-slab reflectionless coating.

```
na=1; nb=1.5; n1=1.38; n2=2.45;
n = [na,n1,n2,nb]; la0 = 550;
r = n2r(n);

c = sqrt((r(1)^2*(1-r(2)*r(3))^2 - (r(2)-r(3))^2)/(4*r(2)*r(3)*(1-r(1)^2)));
k2l2 = acos(c);
G2 = (r(2)+r(3)*exp(-2*j*k2l2))/(1 + r(2)*r(3)*exp(-2*j*k2l2));
k1l1 = (angle(G2) - pi - angle(r(1)))/2;
if k1l1 <0, k1l1 = k1l1 + 2*pi; end

L = [k1l1,k2l2]/2/pi;

la = linspace(400,700,101);
Ga = abs(multidiel(n, L, la/la0)).^2 * 100;
Gb = abs(multidiel([na,n1,nb], 0.25, la/la0)).^2 * 100;
Gc = abs(multidiel([na,sqrt(nb),nb], 0.25, la/la0)).^2 * 100;

plot(la, Ga, la, Gb, la, Gc);
```

The dependence on λ comes through the quantities k_1l_1 and k_2l_2, for example:

$$k_1l_1 = 2\pi\frac{n_1l_1}{\lambda} = 2\pi\frac{0.3294\lambda_0}{\lambda}$$

Essentially the same method is used in Sec. 12.7 to design 2-section series impedance transformers. The MATLAB function `twosect` of that section implements the design. It can be used to obtain the optical lengths of the layers, and in fact, it produces two possible solutions:

$$L_{12} = \text{twosect}(1,\, 1/1.38,\, 1/2.45,\, 1/1.5) = \begin{bmatrix} 0.3294 & 0.0453 \\ 0.1706 & 0.4547 \end{bmatrix}$$

where each row represents a solution, so that $L_1 = n_1l_1/\lambda_0 = 0.1706$ and $L_2 = n_2l_2/\lambda_0 = 0.4547$ is the second solution. The arguments of `twosect` are the inverses of the refractive indices, which are proportional to the characteristic impedances of the four media.

Although this design method meets its design objectives, it results in a narrower bandwidth compared to that of the ideal single-slab case. Varying n_2 has only a minor effect on the shape of the curve. To widen the bandwidth, and at the same time keep the reflection response low, more than two layers must be used.

A simple approach is to fix the optical thicknesses of the films to some prescribed values, such as quarter-wavelengths, and adjust the refractive indices hoping that the required index values come close to realizable ones [594,620]. Fig. 6.2.2 shows the two possible structures: the quarter-quarter two-film case and the quarter-half-quarter three-film case.

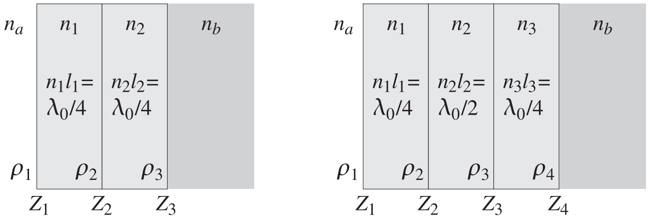


Fig. 6.2.2 Quarter-quarter and quarter-half-quarter antireflection coatings.

The behavior of the two structures is similar at the design wavelength. For the quarter-quarter case, the requirement $Z_1 = \eta_a$ implies:

$$Z_1 = \frac{\eta_1^2}{Z_2} = \frac{\eta_1^2}{\eta_2^2/Z_3} = \frac{\eta_1^2}{\eta_2^2}\eta_b = \eta_a$$

which gives the design condition (see also Example 5.7.1):

$$n_a = \frac{n_1^2}{n_2^2}\,n_b \tag{6.2.3}$$

The optical thicknesses are $n_1 l_1 = n_2 l_2 = \lambda_0/4$. In the quarter-half-quarter case, the half-wavelength layer acts as an absentee layer, that is, $Z_2 = Z_3$, and the resulting design condition is the same:

$$Z_1 = \frac{\eta_1^2}{Z_2} = \frac{\eta_1^2}{Z_3} = \frac{\eta_1^2}{\eta_3^2/Z_4} = \frac{\eta_1^2}{\eta_3^2}\eta_b = \eta_a$$

yielding in the condition:

$$n_a = \frac{n_1^2}{n_3^2}\,n_b \tag{6.2.4}$$

The optical thicknesses are now $n_1 l_1 = n_3 l_3 = \lambda_0/4$ and $n_2 l_2 = \lambda_0/2$. Conditions (6.2.3) and (6.2.4) are the same as far as determining the refractive index of the second quarter-wavelength layer. In the quarter-half-quarter case, the index n_2 of the half-wavelength film is arbitrary.

In the quarter-quarter case, if the first quarter-wave film is magnesium fluoride with $n_1 = 1.38$ and the glass substrate has $n_{\text{glass}} = 1.5$, condition (6.2.3) gives for the index for the second quarter-wave layer:

$$n_2 = \sqrt{\frac{n_1^2 n_b}{n_a}} = \sqrt{\frac{1.38^2 \times 1.50}{1.0}} = 1.69 \tag{6.2.5}$$

The material cerium fluoride (CeF_3) has an index of $n_2 = 1.63$ at $\lambda_0 = 550$ nm and can be used as an approximation to the ideal value of Eq. (6.2.5). Fig. 6.2.3 shows the reflectances $|\Gamma_1|^2$ for the two- and three-layer cases and for the ideal and approximate values of the index of the second quarter-wave layer.

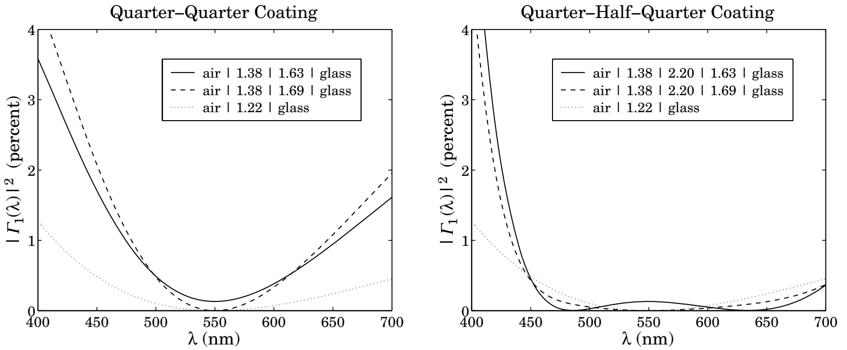


Fig. 6.2.3 Reflectances of the quarter-quarter and quarter-half-quarter cases.

The design wavelength was $\lambda_0 = 550$ nm and the index of the half-wave slab was $n_2 = 2.2$ corresponding to zirconium oxide (ZrO_2). We note that the quarter-half-quarter case achieves a much broader bandwidth over most of the visible spectrum, for either value of the refractive index of the second quarter slab.

The reflectances were computed with the help of the function `multidiel`. The typical MATLAB code was as follows:

```
la0 = 550; la = linspace(400,700,101);

Ga = 100*abs(multidiel([1,1.38,2.2,1.63,1.5], [0.25,0.5,0.25], la/la0)).^2;
Gb = 100*abs(multidiel([1,1.38,2.2,1.69,1.5], [0.25,0.5,0.25], la/la0)).^2;
Gc = 100*abs(multidiel([1,1.22,1.5], 0.25, la/la0)).^2;

plot(la, Ga, la, Gb, la, Gc);
```

These and other methods of designing and manufacturing antireflection coatings for glasses and other substrates can be found in the vast thin-film literature. An incomplete set of references is [592–652]. Some typical materials used in thin-film coatings are given below:

material	n	material	n
cryolite (Na_3AlF_6)	1.35	magnesium fluoride (MgF_2)	1.38
Silicon dioxide SiO_2	1.46	polystyrene	1.60
cerium fluoride (CeF_3)	1.63	lead fluoride (PbF_2)	1.73
Silicon monoxide SiO	1.95	zirconium oxide (ZrO_2)	2.20
zinc sulfide (ZnS)	2.32	titanium dioxide (TiO_2)	2.40
bismuth oxide (Bi_2O_3)	2.45	silicon (Si)	3.50
germanium (Ge)	4.20	tellurium (Te)	4.60

Thin-film coatings have a wide range of applications, such as displays; camera lenses, mirrors, and filters; eyeglasses; coatings for energy-saving lamps and architectural windows; lighting for dental, surgical, and stage environments; heat reflectors for movie projectors; instrumentation, such as interference filters for spectroscopy, beam splitters and mirrors, laser windows, and polarizers; optics of photocopiers and compact disks; optical communications; home appliances, such as heat reflecting oven windows; rear-view mirrors for automobiles.

6.3 Dielectric Mirrors

The main interest in dielectric mirrors is that they have extremely low losses at optical and infrared frequencies, as compared to ordinary metallic mirrors. On the other hand, metallic mirrors reflect over a wider bandwidth than dielectric ones and from all incident angles. However, *omnidirectional dielectric mirrors* are also possible and have recently been constructed [736,737]. The omnidirectional property is discussed in Sec. 8.8. Here, we consider only the normal-incidence case.

A dielectric mirror (also known as a Bragg reflector) consists of identical alternating layers of high and low refractive indices, as shown in Fig. 6.3.1. The optical thicknesses are typically chosen to be quarter-wavelength long, that is, $n_H l_H = n_L l_L = \lambda_0/4$ at some operating wavelength λ_0. The standard arrangement is to have an odd number of layers, with the high index layer being the first and last layer.

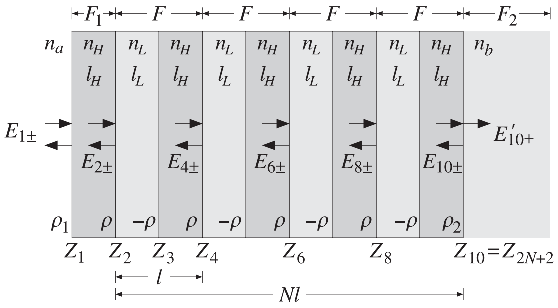


Fig. 6.3.1 Nine-layer dielectric mirror.

Fig. 6.3.1 shows the case of nine layers. If the number of layers is $M = 2N + 1$, the number of interfaces will be $2N + 2$ and the number of media $2N + 3$. After the first layer, we may view the structure as the repetition of N identical bilayers of low and high index. The elementary reflection coefficients alternate in sign as shown in Fig. 6.3.1 and are given by

$$\rho = \frac{n_H - n_L}{n_H + n_L}\,, \quad -\rho = \frac{n_L - n_H}{n_L + n_H}\,, \quad \rho_1 = \frac{n_a - n_H}{n_a + n_H}\,, \quad \rho_2 = \frac{n_H - n_b}{n_H + n_b} \tag{6.3.1}$$

The substrate n_b can be arbitrary, even the same as the incident medium n_a. In that case, $\rho_2 = -\rho_1$. The reflectivity properties of the structure can be understood by propagating the impedances from bilayer to bilayer. For the example of Fig. 6.3.1, we have for the quarter-wavelength case:

$$Z_2 = \frac{\eta_L^2}{Z_3} = \frac{\eta_L^2}{\eta_H^2} Z_4 = \left(\frac{n_H}{n_L}\right)^2 Z_4 = \left(\frac{n_H}{n_L}\right)^4 Z_6 = \left(\frac{n_H}{n_L}\right)^6 Z_8 = \left(\frac{n_H}{n_L}\right)^8 \eta_b$$

Therefore, after each bilayer, the impedance decreases by a factor of $(n_L/n_H)^2$. After N bilayers, we will have:

$$Z_2 = \left(\frac{n_H}{n_L}\right)^{2N} \eta_b \tag{6.3.2}$$

Using $Z_1 = \eta_H^2/Z_2$, we find for the reflection response at λ_0:

$$\Gamma_1 = \frac{Z_1 - \eta_a}{Z_1 + \eta_a} = \frac{1 - \left(\dfrac{n_H}{n_L}\right)^{2N} \dfrac{n_H^2}{n_a n_b}}{1 + \left(\dfrac{n_H}{n_L}\right)^{2N} \dfrac{n_H^2}{n_a n_b}} \tag{6.3.3}$$

It follows that for large N, Γ_1 will tend to -1, that is, 100 % reflection.

Example 6.3.1: For nine layers, $2N + 1 = 9$, or $N = 4$, and $n_H = 2.32$, $n_L = 1.38$, and $n_a = n_b = 1$, we find:

$$\Gamma_1 = \frac{1 - \left(\dfrac{2.32}{1.38}\right)^8 2.32^2}{1 + \left(\dfrac{2.32}{1.38}\right)^8 2.32^2} = -0.9942 \quad \Rightarrow \quad |\Gamma_1|^2 = 98.84 \text{ percent}$$

For $N = 8$, or 17 layers, we have $\Gamma_1 = -0.9999$ and $|\Gamma_1|^2 = 99.98$ percent. If the substrate is glass with $n_b = 1.52$, the reflectances change to $|\Gamma_1|^2 = 98.25$ percent for $N = 4$, and $|\Gamma_1|^2 = 99.97$ percent for $N = 8$. □

To determine the bandwidth around λ_0 for which the structure exhibits high reflectivity, we work with the layer recursions (6.1.2). Because the bilayers are identical, the forward/backward fields at the left of one bilayer are related to those at the left of the next one by a transition matrix F, which is the product of two propagation matrices of the type of Eq. (6.1.2). The repeated application of the matrix F takes us to the right-most layer. For example, in Fig. 6.3.1 we have:

$$\begin{bmatrix} E_{2+} \\ E_{2-} \end{bmatrix} = F \begin{bmatrix} E_{4+} \\ E_{4-} \end{bmatrix} = F^2 \begin{bmatrix} E_{6+} \\ E_{6-} \end{bmatrix} = F^3 \begin{bmatrix} E_{8+} \\ E_{8-} \end{bmatrix} = F^4 \begin{bmatrix} E_{10+} \\ E_{10-} \end{bmatrix}$$

where F is the matrix:

$$F = \frac{1}{1+\rho}\begin{bmatrix} e^{jk_Ll_L} & \rho e^{-jk_Ll_L} \\ \rho e^{jk_Ll_L} & e^{-jk_Ll_L}\end{bmatrix} \frac{1}{1-\rho}\begin{bmatrix} e^{jk_Hl_H} & -\rho e^{-jk_Hl_H} \\ -\rho e^{jk_Hl_H} & e^{-jk_Hl_H}\end{bmatrix} \tag{6.3.4}$$

Defining the phase thicknesses $\delta_H = k_Hl_H$ and $\delta_L = k_Ll_L$, and multiplying the matrix factors out, we obtain the expression for F:

$$F = \frac{1}{1-\rho^2}\begin{bmatrix} e^{j(\delta_H+\delta_L)} - \rho^2 e^{j(\delta_H-\delta_L)} & -2j\rho e^{-j\delta_H}\sin\delta_L \\ 2j\rho e^{j\delta_H}\sin\delta_L & e^{-j(\delta_H+\delta_L)}-\rho^2 e^{-j(\delta_H-\delta_L)}\end{bmatrix} \tag{6.3.5}$$

By an additional transition matrix F_1 we can get to the left of interface-1 and by an additional matching matrix F_2 we pass to the right of the last interface:

$$\begin{bmatrix}E_{1+}\\E_{1-}\end{bmatrix} = F_1\begin{bmatrix}E_{2+}\\E_{2-}\end{bmatrix} = F_1F^4\begin{bmatrix}E_{10+}\\E_{10-}\end{bmatrix} = F_1F^4F_2\begin{bmatrix}E'_{10+}\\0\end{bmatrix}$$

where F_1 and F_2 are:

$$F_1 = \frac{1}{\tau_1}\begin{bmatrix} e^{jk_Hl_H} & \rho_1 e^{-jk_Hl_H} \\ \rho_1 e^{jk_Hl_H} & e^{-jk_Hl_H}\end{bmatrix},\quad F_2=\frac{1}{\tau_2}\begin{bmatrix}1&\rho_2\\ \rho_2&1\end{bmatrix}\tag{6.3.6}$$

where $\tau_1 = 1+\rho_1$, $\tau_2 = 1+\rho_2$, and ρ_1,ρ_2 were defined in Eq. (6.3.1). More generally, for $2N+1$ layers, or N bilayers, we have:

$$\begin{bmatrix}E_{2+}\\E_{2-}\end{bmatrix} = F^N\begin{bmatrix}E_{2N+2,+}\\E_{2N+2,-}\end{bmatrix},\quad \begin{bmatrix}E_{1+}\\E_{1-}\end{bmatrix} = F_1F^NF_2\begin{bmatrix}E'_{2N+2,+}\\0\end{bmatrix}\tag{6.3.7}$$

Thus, the properties of the multilayer structure are essentially determined by the Nth power, F^N, of the bilayer transition matrix F. In turn, the behavior of F^N is determined by the eigenvalue structure of F.

Let $\{\lambda_+,\lambda_-\}$ be the two eigenvalues of F and let V be the eigenvector matrix. Then, the eigenvalue decomposition of F and F^N will be $F = V\Lambda V^{-1}$ and $F^N = V\Lambda^N V^{-1}$, where $\Lambda = \mathrm{diag}\{\lambda_+,\lambda_-\}$. Because F has unit determinant, its two eigenvalues will be inverses of each other, that is, $\lambda_- = 1/\lambda_+$, or, $\lambda_+\lambda_- = 1$.

The eigenvalues $\lambda_\pm$ are either both real-valued or both complex-valued with unit magnitude. We can represent them in the equivalent form:

$$\lambda_+ = e^{jKl},\quad \lambda_- = e^{-jKl}\tag{6.3.8}$$

where l is the length of each bilayer, $l = l_L + l_H$. The quantity K is referred to as the Bloch wavenumber. If the eigenvalues $\lambda_\pm$ are unit-magnitude complex-valued, then K is real. If the eigenvalues are real, then K is pure imaginary, say $K = -j\alpha$, so that $\lambda_\pm = e^{\pm jKl} = e^{\pm\alpha l}$.

The multilayer structure behaves very differently depending on the nature of K. The structure is primarily reflecting if K is imaginary and the eigenvalues $\lambda_\pm$ are real, and it is primarily transmitting if K is real and the eigenvalues are pure phases. To see this, we write Eq. (6.3.7) in the form:

$$\begin{bmatrix}E_{2+}\\E_{2-}\end{bmatrix} = V\Lambda^N V^{-1}\begin{bmatrix}E_{2N+2,+}\\E_{2N+2,-}\end{bmatrix} \quad\Rightarrow\quad V^{-1}\begin{bmatrix}E_{2+}\\E_{2-}\end{bmatrix} = \Lambda^N V^{-1}\begin{bmatrix}E_{2N+2,+}\\E_{2N+2,-}\end{bmatrix},\quad\text{or,}$$

$$\begin{bmatrix}V_{2+}\\V_{2-}\end{bmatrix} = \Lambda^N\begin{bmatrix}V_{2N+2,+}\\V_{2N+2,-}\end{bmatrix}$$

where we defined

$$\begin{bmatrix}V_{2+}\\V_{2-}\end{bmatrix} = V^{-1}\begin{bmatrix}E_{2+}\\E_{2-}\end{bmatrix},\quad \begin{bmatrix}V_{2N+2,+}\\V_{2N+2,-}\end{bmatrix} = V^{-1}\begin{bmatrix}E_{2N+2,+}\\E_{2N+2,-}\end{bmatrix}$$

We have $V_{2+} = \lambda_+^N V_{2N+2,+}$ and $V_{2-} = \lambda_-^N V_{2N+2,-} = \lambda_+^{-N}V_{2N+2,-}$ because Λ^N is diagonal. Thus,

$$V_{2N+2,+} = \lambda_+^{-N}V_{2+} = e^{-jKNl}V_{2+},\quad V_{2N+2,-} = \lambda_+^N V_{2-} = e^{jKNl}V_{2-}\tag{6.3.9}$$

The quantity Nl is recognized as the total length of the bilayer structure, as depicted in Fig. 6.3.1. It follows that if K is real, the factor $\lambda_+^{-N} = e^{-jKNl}$ acts as a propagation phase factor and the fields transmit through the structure.

On the other hand, if K is imaginary, we have $\lambda_+^{-N} = e^{-\alpha Nl}$ and the fields attenuate exponentially as they propagate into the structure. In the limit of large N, the transmitted fields attenuate completely and the structure becomes 100% reflecting. For finite but large N, the structure will be mostly reflecting.

The eigenvalues $\lambda_\pm$ switch from real to complex, as K switches from imaginary to real, for certain frequency or wavenumber bands. The edges of these bands determine the bandwidths over which the structure will act as a mirror.

The eigenvalues are determined from the characteristic polynomial of F, given by the following expression which is valid for any 2×2 matrix:

$$\det(F - \lambda I) = \lambda^2 - (\mathrm{tr}\,F)\lambda + \det F\tag{6.3.10}$$

where I is the 2×2 identity matrix. Because (6.3.5) has unit determinant, the eigenvalues are the solutions of the quadratic equation:

$$\lambda^2 - (\mathrm{tr}\,F)\lambda + 1 = \lambda^2 - 2a\lambda + 1 = 0\tag{6.3.11}$$

where we defined $a = (\mathrm{tr}\,F)/2$. The solutions are:

$$\lambda_\pm = a \pm \sqrt{a^2-1}\tag{6.3.12}$$

where it follows from Eq. (6.3.5) that a is given by:

$$\boxed{a = \frac{1}{2}\,\mathrm{tr}\,F = \frac{\cos(\delta_H+\delta_L) - \rho^2\cos(\delta_H-\delta_L)}{1-\rho^2}}\tag{6.3.13}$$

Using $\lambda_+ = e^{jKl} = a + \sqrt{a^2-1} = a + j\sqrt{1-a^2}$, we also find:

$$\boxed{a=\cos Kl \quad\Rightarrow\quad K=\frac{1}{l}\,\mathrm{acos}(a)} \tag{6.3.14}$$

The sign of the quantity a^2-1 determines whether the eigenvalues are real or complex. The eigenvalues switch from real to complex—equivalently, K switches from imaginary to real—when $a^2=1$, or, $a=\pm 1$. These critical values of K are found from Eq. (6.3.14) to be:

$$K=\frac{1}{l}\,\mathrm{acos}(\pm 1)=\frac{m\pi}{l} \tag{6.3.15}$$

where m is an integer. The lowest value is $K=\pi/l$ and corresponds to $a=-1$ and to $\lambda_+=e^{jKl}=e^{j\pi}=-1$. Thus, we obtain the bandedge condition:

$$a=\frac{\cos(\delta_H+\delta_L)-\rho^2\cos(\delta_H-\delta_L)}{1-\rho^2}=-1$$

It can be manipulated into:

$$\boxed{\cos^2\Big(\frac{\delta_H+\delta_L}{2}\Big)=\rho^2\cos^2\Big(\frac{\delta_H-\delta_L}{2}\Big)} \tag{6.3.16}$$

The dependence on the free-space wavelength λ or frequency $f=c_0/\lambda$ comes through $\delta_H=2\pi(n_Hl_H)/\lambda$ and $\delta_L=2\pi(n_Ll_L)/\lambda$. The solutions of (6.3.16) in λ determine the left and right bandedges of the reflecting regions.

These solutions can be obtained numerically with the help of the MATLAB function `omniband`, discussed in Sec. 8.8. An approximate solution, which is exact in the case of quarter-wave layers, is given below.

If the high and low index layers have equal optical thicknesses, $n_Hl_H=n_Ll_L$, such as when they are quarter-wavelength layers, or when the optical lengths are approximately equal, we can make the approximation $\cos\big((\delta_H-\delta_L)/2\big)=1$. Then, (6.3.16) simplifies into:

$$\cos^2\Big(\frac{\delta_H+\delta_L}{2}\Big)=\rho^2 \tag{6.3.17}$$

with solutions:

$$\cos\Big(\frac{\delta_H+\delta_L}{2}\Big)=\pm\rho \quad\Rightarrow\quad \frac{\delta_H+\delta_L}{2}=\frac{\pi(n_Hl_H+n_Ll_L)}{\lambda}=\mathrm{acos}(\pm\rho)$$

The solutions for the left and right bandedges and the bandwidth in λ are:

$$\lambda_1=\frac{\pi(n_Hl_H+n_Ll_L)}{\mathrm{acos}(-\rho)}\,,\quad \lambda_2=\frac{\pi(n_Hl_H+n_Ll_L)}{\mathrm{acos}(\rho)}\,,\quad \Delta\lambda=\lambda_2-\lambda_1 \tag{6.3.18}$$

Similarly, the left/right bandedges in frequency are $f_1=c_0/\lambda_2$ and $f_2=c_0/\lambda_1$:

$$f_1=c_0\,\frac{\mathrm{acos}(\rho)}{\pi(n_Hl_H+n_Ll_L)}\,,\quad f_2=c_0\,\frac{\mathrm{acos}(-\rho)}{\pi(n_Hl_H+n_Ll_L)} \tag{6.3.19}$$

Noting that $\mathrm{acos}(-\rho)=\pi/2+\mathrm{asin}(\rho)$ and $\mathrm{acos}(\rho)=\pi/2-\mathrm{asin}(\rho)$, the frequency bandwidth can be written in the equivalent forms:

$$\Delta f=f_2-f_1=c_0\,\frac{\mathrm{acos}(-\rho)-\mathrm{acos}(\rho)}{\pi(n_Hl_H+n_Ll_L)}=c_0\,\frac{2\,\mathrm{asin}(\rho)}{\pi(n_Hl_H+n_Ll_L)} \tag{6.3.20}$$

Relative to some desired wavelength $\lambda_0=c_0/f_0$, the normalized bandwidths in wavelength and frequency are:

$$\frac{\Delta\lambda}{\lambda_0}=\frac{\pi(n_Hl_H+n_Ll_L)}{\lambda_0}\left[\frac{1}{\mathrm{acos}(\rho)}-\frac{1}{\mathrm{acos}(-\rho)}\right] \tag{6.3.21}$$

$$\frac{\Delta f}{f_0}=\frac{2\lambda_0\,\mathrm{asin}(\rho)}{\pi(n_Hl_H+n_Ll_L)} \tag{6.3.22}$$

Similarly, the center of the reflecting band $f_c=(f_1+f_2)/2$ is:

$$\frac{f_c}{f_0}=\frac{\lambda_0}{2(n_Hl_H+n_Ll_L)} \tag{6.3.23}$$

If the layers have equal quarter-wave optical lengths at λ_0, that is, $n_Hl_H=n_Ll_L=\lambda_0/4$, then, $f_c=f_0$ and the matrix F takes the simplified form:

$$F=\frac{1}{1-\rho^2}\begin{bmatrix} e^{2j\delta}-\rho^2 & -2j\rho e^{-j\delta}\sin\delta \\ 2j\rho e^{j\delta}\sin\delta & e^{-2j\delta}-\rho^2 \end{bmatrix} \tag{6.3.24}$$

where $\delta=\delta_H=\delta_L=2\pi(n_Hl_H)/\lambda=2\pi(\lambda_0/4)/\lambda=(\pi/2)\lambda_0/\lambda=(\pi/2)f/f_0$. Then, Eqs. (6.3.21) and (6.3.22) simplify into:

$$\frac{\Delta\lambda}{\lambda_0}=\frac{\pi}{2}\left[\frac{1}{\mathrm{acos}(\rho)}-\frac{1}{\mathrm{acos}(-\rho)}\right],\quad \frac{\Delta f}{f_0}=\frac{4}{\pi}\,\mathrm{asin}(\rho) \tag{6.3.25}$$

Example 6.3.2: *Dielectric Mirror With Quarter-Wavelength Layers.* Fig. 6.3.2 shows the reflection response $|\Gamma_1|^2$ as a function of the free-space wavelength λ and as a function of frequency $f=c_0/\lambda$. The high and low indices are $n_H=2.32$ and $n_L=1.38$, corresponding to zinc sulfide (ZnS) and magnesium fluoride. The incident medium is air and the substrate is glass with indices $n_a=1$ and $n_b=1.52$. The left graph depicts the response for the cases of $N=2,4,8$ bilayers, or $2N+1=5,9,17$ layers, as defined in Fig. 6.3.1. The design wavelength at which the layers are quarter-wavelength long is $\lambda_0=500$ nm.

The reflection coefficient is $\rho=0.25$ and the ratio $n_H/n_L=1.68$. The wavelength bandwidth calculated from Eq. (6.3.25) is $\Delta\lambda=168.02$ nm and has been placed on the graph at an arbitrary reflectance level. The left/right bandedges are $\lambda_1=429.73$, $\lambda_2=597.75$ nm. The bandwidth covers most of the visible spectrum. As the number of bilayers N increases, the reflection response becomes flatter within the bandwidth $\Delta\lambda$, and has sharper edges and tends to 100%. The bandwidth $\Delta\lambda$ represents the asymptotic width of the reflecting band.

The right figure depicts the reflection response as a function of frequency f and is plotted in the normalized variable f/f_0. Because the phase thickness of each layer is $\delta=\pi f/2f_0$ and the matrix F is periodic in δ, the mirror behavior of the structure will occur at odd multiples of f_0 (or odd multiples of $\pi/2$ for δ.) As discussed in Sec. 6.6, the structure acts as a sampled system with sampling frequency $f_s=2f_0$, and therefore, $f_0=f_s/2$ plays the role of the Nyquist frequency.

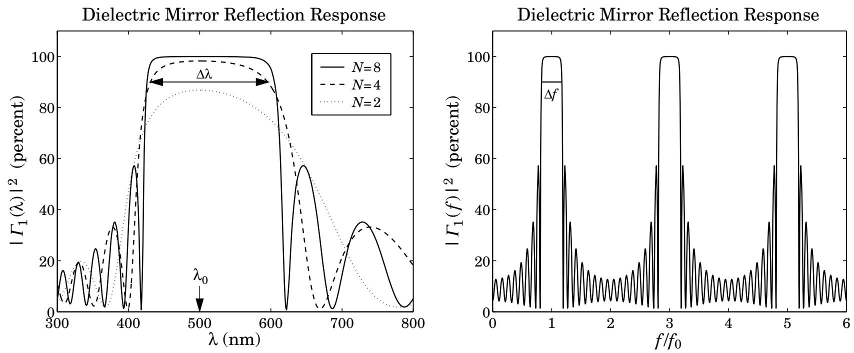


Fig. 6.3.2 Dielectric mirror with quarter-wavelength layers.

The typical MATLAB code used to generate these graphs was:

```
na = 1; nb = 1.52; nH = 2.32; nL = 1.38;      % refractive indices
LH = 0.25; LL = 0.25;                         % optical thicknesses in units of λ0
la0 = 500;                                    % λ0 in units of nm

rho = (nH-nL)/(nH+nL);                        % reflection coefficient ρ

la2 = pi*(LL+LH)*1/acos(rho) * la0;           % right bandedge
la1 = pi*(LL+LH)*1/acos(-rho) * la0;          % left bandedge
Dla = la2-la1;                                % bandwidth

N = 8;                                        % number of bilayers
n = [na, nH, repmat([nL,nH], 1, N), nb];      % indices for the layers A|H(LH)^N|G
L = [LH, repmat([LL,LH], 1, N)];              % lengths of the layers H(LH)^N

la = linspace(300,800,501);                   % plotting range is 300 ≤ λ ≤ 800 nm
Gla = 100*abs(multidiel(n,L,la/la0)).^2;      % reflectance as a function of λ
figure; plot(la,Gla);

f = linspace(0,6,1201);                       % frequency plot over 0 ≤ f ≤ 6f0
Gf = 100*abs(multidiel(n,L,1./f)).^2;         % reflectance as a function of f
figure; plot(f,Gf);
```

Note that the function `repmat` replicates the LH bilayer N times. The frequency graph shows only the case of $N = 8$. The bandwidth Δf, calculated from (6.3.25), has been placed on the graph. The maximum reflectance (evaluated at odd multiples of f_0) is equal to 99.97%. ☐

Example 6.3.3: *Dielectric Mirror with Unequal-Length Layers.* Fig. 6.3.3 shows the reflection response of a mirror having unequal optical lengths for the high and low index films.

The parameters of this example correspond very closely to the recently constructed omnidirectional dielectric mirror [736], which was designed to be a mirror over the infrared band of 10–15 μm. The number of layers is nine and the number of bilayers, $N = 4$. The indices of refraction are $n_H = 4.6$ and $n_L = 1.6$ corresponding to Tellurium and Polystyrene.

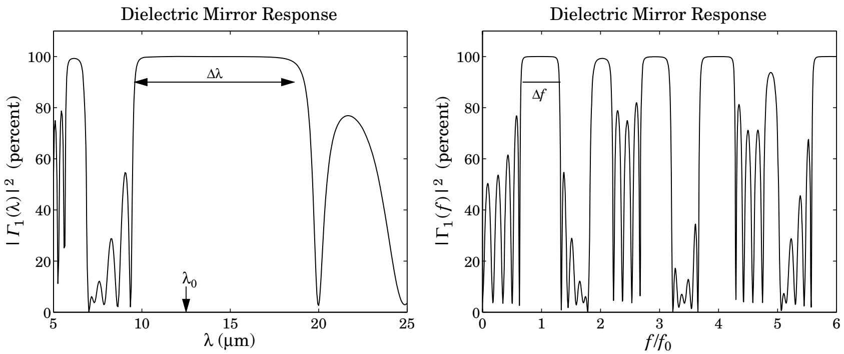


Fig. 6.3.3 Dielectric mirror with unequal optical thicknesses.

Their ratio is $n_H/n_L = 2.875$ and the reflection coefficient, $\rho = 0.48$. The incident medium and substrate are air and NaCl ($n = 1.48$.)

The center wavelength is taken to be at the middle of the 10–15 μm band, that is, $\lambda_0 = 12.5$ μm. The lengths of the layers are $l_H = 0.8$ and $l_L = 1.65$ μm, resulting in the optical lengths (relative to λ_0) $n_H l_H = 0.2944\lambda_0$ and $n_L l_L = 0.2112\lambda_0$. The wavelength bandwidth, calculated from Eq. (6.3.21), is $\Delta\lambda = 9.07$ μm. The typical MATLAB code for generating the figures of this example was as follows:

```
la0 = 12.5;
na = 1; nb = 1.48;                            % NaCl substrate
nH = 4.6; nL = 1.6;                           % Te and PS
lH = 0.8; lL = 1.65;                          % physical lengths lH,lL
LH = nH*lH/la0, LL = nL*lL/la0;               % optical lengths in units of λ0

rho = (nH-nL)/(nH+nL);                        % reflection coefficient ρ

la2 = pi*(LL+LH)*1/acos(rho) * la0;           % right bandedge
la1 = pi*(LL+LH)*1/acos(-rho) * la0;          % left bandedge
Dla = la2-la1;                                % bandwidth

la = linspace(5,25,401);                      % equally-spaced wavelengths

N = 4;
n = [na, nH, repmat([nL,nH], 1, N), nb];      % refractive indices of all media
L = [LH, repmat([LL,LH], 1, N)];              % optical lengths of the slabs
G = 100 * abs(multidiel(n,L,la/la0)).^2;      % reflectance

plot(la,G);
```

The bandwidth $\Delta\lambda$ shown on the graph is wider than that of the omnidirectional mirror presented in [736], because our analysis assumes normal incidence only. The condition for omnidirectional reflectivity for both TE and TM modes causes the bandwidth to narrow by about half of what is shown in the figure. The reflectance as a function of frequency is no longer periodic at odd multiples of f_0 because the layers have lengths that are not equal to $\lambda_0/4$. The omnidirectional case is discussed in Example 8.8.3.

The maximum reflectivity achieved within the mirror bandwidth is 99.99%, which is better than that of the previous example with 17 layers. This can be explained because the ratio n_H/n_L is much larger here. □

Although the reflectances in the previous two examples were computed with the help of the MATLAB function `multidiel`, it is possible to derive closed-form expressions for Γ_1 that are valid for any number of bilayers N. Applying Eq. (6.1.3) to interface-1 and interface-2, we have:

$$\Gamma_1 = \frac{\rho_1 + e^{-2j\delta_H}\Gamma_2}{1 + \rho_1 e^{-2j\delta_H}\Gamma_2} \tag{6.3.26}$$

where $\Gamma_2 = E_{2-}/E_{2+}$, which can be computed from the matrix equation (6.3.7). Thus, we need to obtain a closed-form expression for Γ_2.

It is a general property of any 2×2 unimodular matrix F that its Nth power can be obtained from the following simple formula, which involves the Nth powers of its eigenvalues $\lambda_\pm$:[†]

$$F^N = \left(\frac{\lambda_+^N - \lambda_-^N}{\lambda_+ - \lambda_-}\right)F - \left(\frac{\lambda_+^{N-1} - \lambda_-^{N-1}}{\lambda_+ - \lambda_-}\right)I = W_N F - W_{N-1} I \tag{6.3.27}$$

where $W_N = (\lambda_+^N - \lambda_-^N)/(\lambda_+ - \lambda_-)$. To prove it, we note that the formula holds as a simple identity when F is replaced by its diagonal version $\Lambda = \mathrm{diag}\{\lambda_+, \lambda_-\}$:

$$\Lambda^N = \left(\frac{\lambda_+^N - \lambda_-^N}{\lambda_+ - \lambda_-}\right)\Lambda - \left(\frac{\lambda_+^{N-1} - \lambda_-^{N-1}}{\lambda_+ - \lambda_-}\right)I \tag{6.3.28}$$

Eq. (6.3.27) then follows by multiplying (6.3.28) from left and right by the eigenvector matrix V and using $F = V\Lambda V^{-1}$ and $F^N = V\Lambda^N V^{-1}$. Defining the matrix elements of F and F^N by

$$F = \begin{bmatrix} A & B \\ B^* & A^* \end{bmatrix}, \quad F^N = \begin{bmatrix} A_N & B_N \\ B_N^* & A_N^* \end{bmatrix}, \tag{6.3.29}$$

it follows from (6.3.27) that:

$$A_N = AW_N - W_{N-1}, \quad B_N = BW_N \tag{6.3.30}$$

where we defined:

$$A = \frac{e^{j(\delta_H+\delta_L)} - \rho^2 e^{j(\delta_H-\delta_L)}}{1-\rho^2}, \quad B = -\frac{2j\rho e^{-j\delta_H}\sin\delta_L}{1-\rho^2} \tag{6.3.31}$$

Because F and F^N are unimodular, their matrix elements satisfy the conditions:

$$|A|^2 - |B|^2 = 1, \quad |A_N|^2 - |B_N|^2 = 1 \tag{6.3.32}$$

The first follows directly from the definition (6.3.29), and the second can be verified easily. It follows now that the product $F^N F_2$ in Eq. (6.3.7) is:

[†] The coefficients W_N are related to the Chebyshev polynomials of the second kind $U_m(x)$ through $W_N = U_{N-1}(a) = \sin(N\,\mathrm{acos}(a))/\sqrt{1-a^2} = \sin(NKl)/\sin(Kl)$.

$$F^N F_2 = \frac{1}{\tau_2}\begin{bmatrix} A_N + \rho_2 B_N & B_N + \rho_2 A_N \\ B_N^* + \rho_2 A_N^* & A_N^* + \rho_2 B_N^* \end{bmatrix}$$

Therefore, the desired closed-form expression for the reflection coefficient Γ_2 is:

$$\Gamma_2 = \frac{B_N^* + \rho_2 A_N^*}{A_N + \rho_2 B_N} = \frac{B^* W_N + \rho_2(A^* W_N - W_{N-1})}{AW_N - W_{N-1} + \rho_2 B W_N} \tag{6.3.33}$$

Suppose now that $a^2 < 1$ and the eigenvalues are pure phases. Then, W_N are oscillatory as functions of the wavelength λ or frequency f and the structure will transmit.

On the other hand, if f lies in the mirror bands, so that $a^2 > 1$, then the eigenvalues will be real with $|\lambda_+| > 1$ and $|\lambda_-| < 1$. In the limit of large N, W_N and W_{N-1} will behave like:

$$W_N \to \frac{\lambda_+^N}{\lambda_+ - \lambda_-}, \quad W_{N-1} \to \frac{\lambda_+^{N-1}}{\lambda_+ - \lambda_-}$$

In this limit, the reflection coefficient Γ_2 becomes:

$$\Gamma_2 \to \frac{B^* + \rho_2(A^* - \lambda_+^{-1})}{A - \lambda_+^{-1} + \rho_2 B} \tag{6.3.34}$$

where we canceled some common diverging factors from all terms. Using conditions (6.3.32) and the eigenvalue equation (6.3.11), and recognizing that $\mathrm{Re}(A) = a$, it can be shown that this asymptotic limit of Γ_2 is unimodular, $|\Gamma_2| = 1$, regardless of the value of ρ_2.

This immediately implies that Γ_1 given by Eq. (6.3.26) will also be unimodular, $|\Gamma_1| = 1$, regardless of the value of ρ_1. In other words, the structure tends to become a perfect mirror as the number of bilayers increases.

Next, we discuss some variations on dielectric mirrors that result in (a) multiband mirrors and (b) longpass and shortpass filters that pass long or short wavelengths, in analogy with lowpass and highpass filters that pass low or high frequencies.

Example 6.3.4: *Multiband Reflectors.* The quarter-wave stack of bilayers of Example 6.3.2 can be denoted compactly as $AH(LH)^8G$ (for the case $N = 8$), meaning 'air', followed by a "high-index" quarter-wave layer , followed by four "low/high" bilayers, followed by the "glass" substrate.

Similarly, Example 6.3.3 can be denoted by $A(1.18H)(0.85L\,1.18H)^4G$, where the layer optical lengths have been expressed in units of $\lambda_0/4$, that is, $n_L l_L = 0.85(\lambda_0/4)$ and $n_H l_H = 1.18(\lambda_0/4)$.

Another possibility for a periodic bilayer structure is to replace one or both of the L or H layers by integral multiples thereof [596]. Fig. 6.3.4 shows two such examples. In the first, each H layer has been replaced by a half-wave layer, that is, two quarter-wave layers $2H$, so that the total structure is $A(2H)(L\,2H)^8G$, where n_a,n_b,n_H,n_L are the same as in Example 6.3.2. In the second case, each H has been replaced by a three-quarter-wave layer, resulting in $A(3H)(L\,3H)^8G$.

The mirror peaks at odd multiples of f_0 of Example 6.3.2 get split into two or three peaks each. □

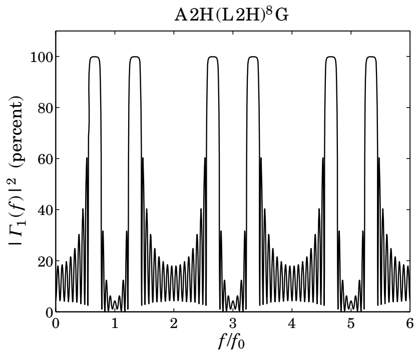


Fig. 6.3.4 Dielectric mirrors with split bands.

Example 6.3.5: *Shortpass and Longpass Filters.* By adding an eighth-wave low-index layer, that is, a $(0.5L)$, at both ends of Example 6.3.2, we can decrease the reflectivity of the short wavelengths. Thus, the stack $AH(LH)^8G$ is replaced by $A(0.5L)H(LH)^8(0.5L)G$.

For example, suppose we wish to have high reflectivity over the $[600, 700]$ nm range and low reflectivity below 500 nm. The left graph in Fig. 6.3.5 shows the resulting reflectance with the design wavelength chosen to be $\lambda_0 = 650$ nm. The parameters n_a, n_b, n_H, n_L are the same as in Example 6.3.2

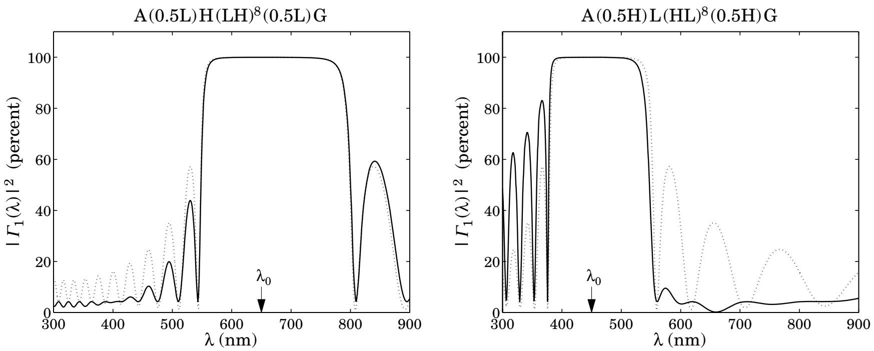


Fig. 6.3.5 Short- and long-pass wavelength filters.

The right graph of Fig. 6.3.5 shows the stack $A(0.5H)L(HL)^8(0.5H)G$ obtained from the previous case by *interchanging* the roles of H and L. Now, the resulting reflectance is low for the higher wavelengths. The design wavelength was chosen to be $\lambda_0 = 450$ nm. It can be seen from the graph that the reflectance is high within the band $[400, 500]$ nm and low above 600 nm.

Superimposed on both graphs is the reflectance of the original $AH(LH)^8G$ stack centered at the corresponding λ_0 (dotted curves.)

Both of these examples can also be thought of as the periodic repetition of a symmetric triple layer of the form $A(BCB)^NG$. Indeed, we have the equivalences:

$$A(0.5L)H(LH)^8(0.5L)G = A(0.5L\,H\,0.5L)^9G$$

$$A(0.5H)L(HL)^8(0.5H)G = A(0.5H\,L\,0.5H)^9G$$

The symmetric triple combination BCB can be replaced by an *equivalent single layer*, which facilitates the analysis of such structures [594,622–624,626]. □

6.4 Propagation Bandgaps

There is a certain analogy between the electronic energy bands of solid state materials arising from the periodicity of the crystal structure and the frequency bands of dielectric mirrors arising from the periodicity of the bilayers. The high-reflectance bands play the role of the forbidden energy bands (in the sense that waves cannot propagate through the structure in these bands.) Such periodic dielectric structures have been termed *photonic crystals* and have given rise to the new field of *photonic bandgap structures*, which has grown rapidly over the past ten years with a large number of potential novel applications [720–746].

Propagation bandgaps arise in any wave propagation problem in a medium with periodic structure [713–719]. Waveguides and transmission lines that are periodically loaded with ridges or shunt impedances, are examples of such media [843–847].

Fiber Bragg gratings, obtained by periodically modulating the refractive index of the core (or the cladding) of a finite portion of a fiber, exhibit high reflectance bands [747–767]. Quarter-wave phase-shifted fiber Bragg gratings (discussed in the next section) act as narrow-band transmission filters and can be used in wavelength multiplexed communications systems.

Other applications of periodic structures with bandgaps arise in structural engineering for the control of vibration transmission and stress [768–770], in acoustics for the control of sound transmission through structures [771–776], and in the construction of laser resonators and periodic lens systems [848,849]. A nice review of wave propagation in periodic structures can be found in [714].

6.5 Narrow-Band Transmission Filters

The reflection bands of a dielectric mirror arise from the N-fold periodic replication of high/low index layers of the type $(HL)^N$, where H, L can have arbitrary lengths. Here, we will assume that they are quarter-wavelength layers at the design wavelength λ_0.

A quarter-wave phase-shifted multilayer structure is obtained by doubling $(HL)^N$ to $(HL)^N(HL)^N$ and then inserting a quarter-wave layer L between the two groups, resulting in $(HL)^NL(HL)^N$. We are going to refer to such a structure as a *Fabry-Perot resonator* (FPR)—it can also be called a *quarter-wave phase-shifted Bragg grating.*

An FPR behaves like a single L-layer at the design wavelength λ_0. Indeed, noting that at λ_0 the combinations LL and HH are half-wave or absentee layers and can be deleted, we obtain the successive reductions:

$$\begin{aligned}(HL)^N L(HL)^N &\to (HL)^{N-1}HLLHL(HL)^{N-1}\\ &\to (HL)^{N-1}HHL(HL)^{N-1}\\ &\to (HL)^{N-1}L(HL)^{N-1}\end{aligned}$$

Thus, the number of the HL layers can be successively reduced, eventually resulting in the equivalent layer L (at λ_0):

$$(HL)^N L(HL)^N \to (HL)^{N-1}L(HL)^{N-1} \to (HL)^{N-2}L(HL)^{N-2} \to \cdots \to L$$

Adding another L-layer on the right, the structure $(HL)^N L(HL)^N L$ will act as $2L$, that is, a half-wave absentee layer at λ_0. If such a structure is sandwiched between the same substrate material, say glass, then it will act as an absentee layer, opening up a narrow transmission window at λ_0, in the middle of its reflecting band.

Without the quarter-wave layers L present, the structures $G|(HL)^N(HL)^N|G$ and $G|(HL)^N|G$ act as mirrors,[†] but with the quarter-wave layers present, the structure $G|(HL)^N L(HL)^N L|G$ acts as a narrow transmission filter, with the transmission bandwidth becoming narrower as N increases.

By repeating the FPR $(HL)^N L(HL)^N$ several times and using possibly different lengths N, it is possible to design a very narrow transmission band centered at λ_0 having a flat passband and very sharp edges.

Thus, we arrive at a whole family of designs, where starting with an ordinary dielectric mirror, we may replace it with one, two, three, four, and so on, FPRs:

$$\begin{aligned}&0.\quad G|(HL)^{N_1}|G\\ &1.\quad G|(HL)^{N_1}L(HL)^{N_1}|L|G\\ &2.\quad G|(HL)^{N_1}L(HL)^{N_1}|(HL)^{N_2}L(HL)^{N_2}|G\\ &3.\quad G|(HL)^{N_1}L(HL)^{N_1}|(HL)^{N_2}L(HL)^{N_2}|(HL)^{N_3}L(HL)^{N_3}|L|G\\ &4.\quad G|(HL)^{N_1}L(HL)^{N_1}|(HL)^{N_2}L(HL)^{N_2}|(HL)^{N_3}L(HL)^{N_3}|(HL)^{N_4}L(HL)^{N_4}|G\end{aligned} \tag{6.5.1}$$

Note that when an odd number of FPRs $(HL)^N L(HL)^N$ are used, an extra L layer must be added at the end to make the overall structure absentee. For an even number of FPRs, this is not necessary.

Such filter designs have been used in thin-film applications [597–603] and in fiber Bragg gratings, for example, as demultiplexers for WDM systems and for generating very-narrow-bandwidth laser sources (typically at $\lambda_0 = 1550$ nm) with distributed feedback lasers [757–767]. We discuss fiber Bragg gratings in Sec. 11.4.

In a Fabry-Perot interferometer, the quarter-wave layer L sandwiched between the mirrors $(HL)^N$ is called a “spacer” or a “cavity” and can be replaced by any odd multiple of quarter-wave layers, for example, $(HL)^N(5L)(HL)^N$.

[†] G denotes the glass substrate.

Several variations of FPR filters are possible, such as interchanging the role of H and L, or using symmetric structures. For example, using eighth-wave layers $L/2$, the following symmetric multilayer structure will also act like as a single L at λ_0:

$$\left(\frac{L}{2}H\frac{L}{2}\right)^N L\left(\frac{L}{2}H\frac{L}{2}\right)^N$$

To create an absentee structure, we may sandwich this between two $L/2$ layers:

$$\frac{L}{2}\left(\frac{L}{2}H\frac{L}{2}\right)^N L\left(\frac{L}{2}H\frac{L}{2}\right)^N\frac{L}{2}$$

This can be seen to be equivalent to $(HL)^N(2L)(LH)^N$, which is absentee at λ_0. This equivalence follows from the identities:

$$\begin{aligned}\frac{L}{2}\left(\frac{L}{2}H\frac{L}{2}\right)^N &\equiv (LH)^N\frac{L}{2}\\ \left(\frac{L}{2}H\frac{L}{2}\right)^N\frac{L}{2} &\equiv \frac{L}{2}(HL)^N\end{aligned} \tag{6.5.2}$$

Example 6.5.1: *Transmission Filter Design with One FPR.* This example illustrates the basic transmission properties of FPR filters. We choose parameters that might closely emulate the case of a fiber Bragg grating for WDM applications. The refractive indices of the left and right substrates and the layers were: $n_a = n_b = 1.52$, $n_L = 1.4$, and $n_H = 2.1$. The design wavelength at which the layers are quarter wavelength is taken to be the standard laser source $\lambda_0 = 1550$ nm.

First, we compare the cases of a dielectric mirror $(HL)^N$ and its phase-shifted version using a single FPR (cases 0 and 1 in Eq. (6.5.1)), with number of layers $N_1 = 6$. Fig. 6.5.1 shows the transmittance, that is, the quantity $(1-|\Gamma_1(\lambda)|^2)$ plotted over the range $1200 \le \lambda \le 2000$ nm.

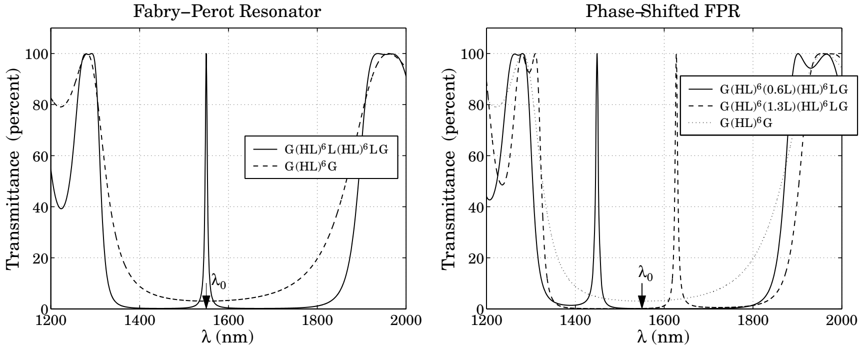


Fig. 6.5.1 Narrowband FPR transmission filters.

We observe that the mirror (case 0) has a suppressed transmittance over the entire reflecting band, whereas the FPR filter (case 1) has a narrow peak at λ_0. The asymptotic edges of the reflecting band are calculated from Eq. (6.3.18) to be $\lambda_1 = 1373.9$ nm and $\lambda_2 = 1777.9$ nm, resulting in a width of $\Delta\lambda = 404$ nm. The MATLAB code used to generated the left graph was:

```
na = 1.52; nb = 1.52; nH = 2.1; nL = 1.4;
LH = 0.25; LL = 0.25;                                  % optical thicknesses

la0 = 1550;
la = linspace(1200, 2000, 8001);                       % 1200 ≤ λ ≤ 2000 nm

N1 = 6;

n1 = repmat([nH,nL],1,N1);
L1 = repmat([LH,LL],1,N1);
n = [na, n1, nb];
L = L1;
G0 = 100*(1 - abs(multidiel(n,L,la/la0)).^2);             % no phase shift

n1 = [repmat([nH,nL],1,N1), nL, repmat([nH,nL],1,N1)];
L1 = [repmat([LH,LL],1,N1), LL, repmat([LH,LL],1,N1)];
n = [na, n1, nL, nb];
L = [L1, LL];
G1 = 100*(1 - abs(multidiel(n,L,la/la0)).^2);             % one phase shift

plot(la,G1,la,G0);
```

The location of the peak can be shifted by making the phase-shift different from $\lambda/4$. This can be accomplished by changing the optical thickness of the middle *L*-layer to some other value. The right graph of Fig. 6.5.1 shows the two cases where that length was chosen to be $n_L l_L = (0.6)\lambda_0/4$ and $(1.3)\lambda_0/4$, corresponding to phase shifts of 54° and 117°. □

Example 6.5.2: *Transmission Filter Design with Two FPRs.* Fig. 6.5.2 shows the transmittance of a grating with two FPRs (case 2 of Eq. (6.5.1)). The number of bilayers were $N_1 = N_2 = 8$ in the first design, and $N_1 = N_2 = 9$ in the second.

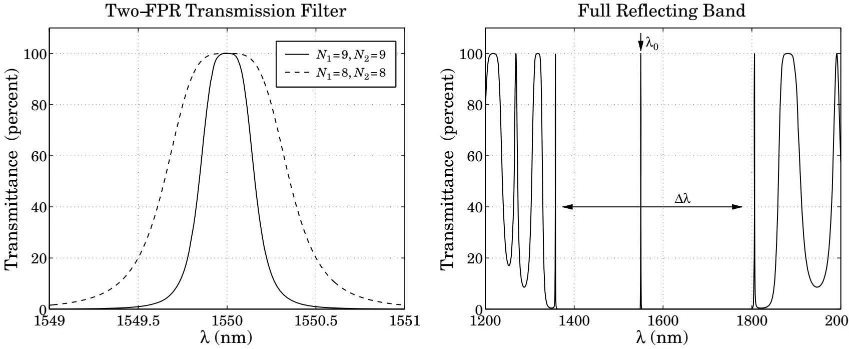


Fig. 6.5.2 Narrow-band transmission filter made with two FPRs.

The resulting transmittance bands are extremely narrow. The plotting scale is only from 1549 nm to 1551 nm. To see these bands in the context of the reflectance band, the transmittance (for $N_1 = N_2 = 8$) is plotted on the right graph over the range [1200, 2000] nm, which includes the full reflectance band of [1373.9, 1777.9] nm.

Using two FPRs has the effect of narrowing the transmittance band and making it somewhat flatter at its top. □

Example 6.5.3: *Transmission Filter Design with Three and Four FPRs.* Fig. 6.5.3 shows the transmittance of a grating with three FPRs (case 3 of Eq. (6.5.1)). A symmetric arrangement of FPRs was chosen such that $N_3 = N_1$.

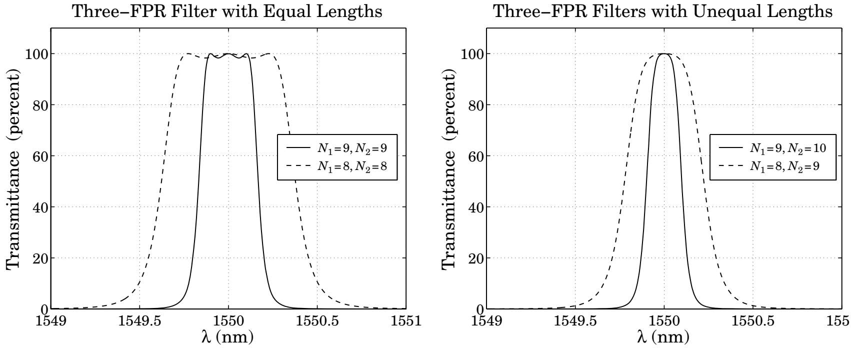


Fig. 6.5.3 Transmission filters with three FPRs of equal and unequal lengths.

The left graph shows the transmittances of the two design cases $N_1 = N_2 = N_3 = 8$ and $N_1 = N_2 = N_3 = 9$, so that all the FPRs have the same lengths. The transmission band is now flatter but exhibits some ripples. To get rid of the ripples, the length of the middle FPR is slightly increased. The right graph shows the case $N_1 = N_3 = 8$ and $N_2 = 9$, and the case $N_1 = N_3 = 9$ and $N_2 = 10$.

Fig. 6.5.4 shows the case of four FPRs (case 4 in Eq. (6.5.1).) Again, a symmetric arrangement was chosen with $N_1 = N_4$ and $N_2 = N_3$.

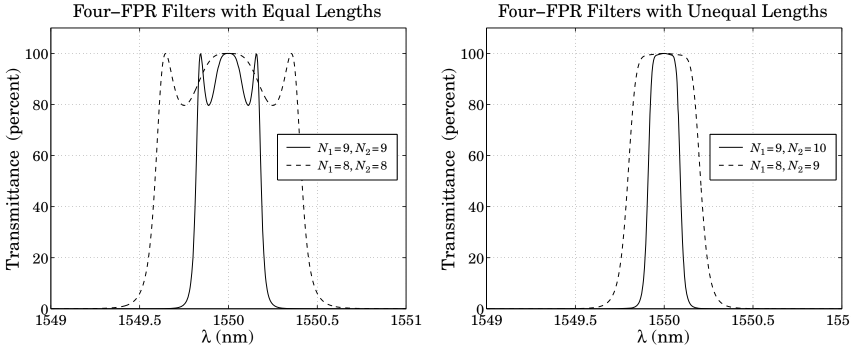


Fig. 6.5.4 Transmission filters with four FPRs of equal and unequal lengths.

The left graph shows the two cases of equal lengths $N_1 = N_2 = N_3 = N_4 = 8$ and $N_1 = N_2 = N_3 = N_4 = 9$. The right graphs shows the case $N_1 = N_4 = 8$ and $N_2 = N_4 = 9$, and the case $N_1 = N_4 = 9$ and $N_2 = N_3 = 10$. We notice again that the equal length cases exhibit ripples, but increasing the length of the middle FPRs tends to eliminate them. The typical MATLAB code for generating the case $N_1 = N_4 = 9$ and $N_2 = N_3 = 10$ was as follows:

```
na = 1.52; nb = 1.52; nH = 2.1; nL = 1.4;
LH = 0.25; LL = 0.25;
la0 = 1550;
la = linspace(1549, 1551, 501);

N1 = 9; N2 = 10; N3 = N2; N4 = N1;

n1 = [repmat([nH,nL],1,N1), nL, repmat([nH,nL],1,N1)];
n2 = [repmat([nH,nL],1,N2), nL, repmat([nH,nL],1,N2)];
n3 = [repmat([nH,nL],1,N3), nL, repmat([nH,nL],1,N3)];
n4 = [repmat([nH,nL],1,N4), nL, repmat([nH,nL],1,N4)];
L1 = [repmat([LH,LL],1,N1), LL, repmat([LH,LL],1,N1)];
L2 = [repmat([LH,LL],1,N2), LL, repmat([LH,LL],1,N2)];
L3 = [repmat([LH,LL],1,N3), LL, repmat([LH,LL],1,N3)];
L4 = [repmat([LH,LL],1,N4), LL, repmat([LH,LL],1,N4)];

n = [na, n1, n2, n3, n4, nb];
L = [L1, L2, L3, L4];

G = 100*(1 - abs(multidiel(n,L,la/la0)).^2);
plot(la,G);
```

The resulting transmittance band is fairly flat with a bandwidth of approximately 0.15 nm, as would be appropriate for dense WDM systems. The second design case with $N_1 = 8$ and $N_2 = 9$ has a bandwidth of about 0.3 nm.

The effect of the relative lengths N_1, N_2 on the shape of the transmittance band has been studied in [763–765]. The equivalence of the low/high multilayer dielectric structures to coupled-mode models of fiber Bragg gratings has been discussed in [754]. ☐

6.6 Equal Travel-Time Multilayer Structures

Here, we discuss the specialized, but useful, case of a multilayer structure whose layers have equal optical thicknesses, or equivalently, equal travel-time delays, as for example in the case of quarter-wavelength layers. Our discussion is based on [792] and on [799,800].

Fig. 6.6.1 depicts such a structure consisting of M layers. The media to the left and right are η_a and η_b and the reflection coefficients ρ_i at the $M+1$ interfaces are as in Eq. (6.1.1). We will discuss the general case when there are incident fields from both the left and right media.

Let T_s denote the common two-way travel-time delay, so that,

$$\frac{2n_1l_1}{c_0} = \frac{2n_2l_2}{c_0} = \cdots = \frac{2n_Ml_M}{c_0} = T_s \tag{6.6.1}$$

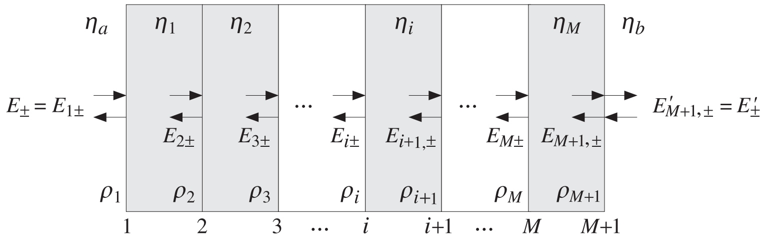


Fig. 6.6.1 Equal travel-time multilayer structure.

Then, all layers have a common phase thickness, that is, for $i = 1, 2, \dots, M$:

$$\delta = k_il_i = \frac{\omega n_il_i}{c_0} = \frac{1}{2}\omega T_s \tag{6.6.2}$$

where we wrote $k_i = \omega/c_i = \omega n_i/c_0$. The layer recursions (6.1.2)–(6.1.5) simplify considerably in this case. These recursions and other properties of the structure can be described using DSP language.

Because the layers have a common roundtrip time delay T_s, the overall structure will act as a sampled system with sampling period T_s and sampling frequency $f_s = 1/T_s$. The corresponding “Nyquist frequency”, $f_0 = f_s/2$, plays a special role. The phase thickness δ can be expressed in terms of f and f_0 as follows:

$$\delta = \frac{1}{2}\omega T_s = \frac{1}{2}2\pi f\frac{1}{f_s} = \pi\frac{f}{f_s} = \frac{\pi}{2}\frac{f}{f_0}$$

Therefore, at $f = f_0$ (and odd multiples thereof), the phase thickness will be $\pi/2 = (2\pi)/4$, that is, the structure will act as quarter-wave layers. Defining the z-domain variable:

$$\boxed{z = e^{2j\delta} = e^{j\omega T_s} = e^{2jk_il_i}} \tag{6.6.3}$$

we write Eq. (6.1.2) in the form:

$$\begin{bmatrix} E_{i+} \\ E_{i-} \end{bmatrix} = \frac{z^{1/2}}{\tau_i}\begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1} \end{bmatrix}\begin{bmatrix} E_{i+1,+} \\ E_{i+1,-} \end{bmatrix}, \quad i = M, M-1, \dots, 1 \tag{6.6.4}$$

We may rewrite it compactly as:

$$\mathbf{E}_i(z) = F_i(z)\mathbf{E}_{i+1}(z) \tag{6.6.5}$$

where we defined:

$$F_i(z) = \frac{z^{1/2}}{\tau_i}\begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1} \end{bmatrix}, \qquad \mathbf{E}_i(z) = \begin{bmatrix} E_{i+}(z) \\ E_{i-}(z) \end{bmatrix} \tag{6.6.6}$$

The transition matrix $F_i(z)$ has two interesting properties. Defining the complex conjugate matrix $\bar{F}_i(z) = F_i(z^{-1})$, we have:

$$\begin{aligned}\bar{F}_i(z)^T J_3 F_i(z) &= \frac{1-\rho_i}{1+\rho_i}J_3 = \frac{\eta_{i-1}}{\eta_i}J_3 \\ \bar{F}_i(z) &= J_1 F_i(z) J_1\end{aligned} \tag{6.6.7}$$

where J_1, J_3 are the 2×2 matrices:[†]

$$J_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad J_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{6.6.8}$$

In proving Eq. (6.6.7), we used the result $(1-\rho_i^2)/\tau_i^2 = (1-\rho_i)/(1+\rho_i) = \eta_{i-1}/\eta_i = n_i/n_{i-1}$. The first of Eqs. (6.6.7) implies *energy conservation*, that is, the energy flux into medium i is equal to the energy flux into medium $i+1$, or,

$$\frac{1}{2\eta_{i-1}}\left(\bar{E}_{i+}E_{i+} - \bar{E}_{i-}E_{i-}\right) = \frac{1}{2\eta_i}\left(\bar{E}_{i+1,+}E_{i+1,+} - \bar{E}_{i+1,-}E_{i+1,-}\right) \tag{6.6.9}$$

This can be expressed compactly in the form:

$$\bar{\mathbf{E}}_i^T J_3 \mathbf{E}_i = \frac{\eta_{i-1}}{\eta_i}\bar{\mathbf{E}}_{i+1}^T J_3 \mathbf{E}_{i+1}$$

which follows from Eq. (6.6.7):

$$\bar{\mathbf{E}}_i^T J_3 \mathbf{E}_i = \bar{\mathbf{E}}_{i+1}^T \bar{F}_i^T J_3 F_i\, \mathbf{E}_{i+1} = \frac{\eta_{i-1}}{\eta_i}\bar{\mathbf{E}}_{i+1}^T J_3 \mathbf{E}_{i+1}$$

The second of Eqs. (6.6.7) expresses *time-reversal invariance* and allows the construction of a second, linearly independent, solution of the recursions (6.6.5):

$$\hat{\mathbf{E}}_i = J_1\bar{\mathbf{E}}_i = \begin{bmatrix} \bar{E}_{i-} \\ \bar{E}_{i+} \end{bmatrix} = J_1\bar{F}_i(z)\bar{\mathbf{E}}_{i+1} = F_i(z)J_1\bar{\mathbf{E}}_{i+1} = F_i(z)\hat{\mathbf{E}}_{i+1}$$

The recursions (6.6.5) may be iterated now to the rightmost interface. By an additional boundary match, we may pass to the right of interface $M+1$:

$$\mathbf{E}_i = F_i(z)F_{i+1}(z)\cdots F_M(z)F_{M+1}\mathbf{E}'_{M+1}$$

where we defined the last transition matrix as

$$F_{M+1} = \frac{1}{\tau_{M+1}}\begin{bmatrix} 1 & \rho_{M+1} \\ \rho_{M+1} & 1 \end{bmatrix} \tag{6.6.10}$$

More explicitly, we have:

$$\begin{aligned}\begin{bmatrix} E_{i+} \\ E_{i-} \end{bmatrix} &= \frac{z^{(M+1-i)/2}}{\nu_i}\begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1} \end{bmatrix}\begin{bmatrix} 1 & \rho_{i+1} z^{-1} \\ \rho_{i+1} & z^{-1} \end{bmatrix}\cdots \\ &\cdots\begin{bmatrix} 1 & \rho_M z^{-1} \\ \rho_M & z^{-1} \end{bmatrix}\begin{bmatrix} 1 & \rho_{M+1} \\ \rho_{M+1} & 1 \end{bmatrix}\begin{bmatrix} E'_{M+1,+} \\ E'_{M+1,-} \end{bmatrix}\end{aligned} \tag{6.6.11}$$

[†] They are recognized as two of the three Pauli spin matrices.

where we defined $\nu_i = \tau_i\tau_{i+1}\cdots\tau_M\tau_{M+1}$. We introduce the following definition for the product of these matrices:

$$\begin{bmatrix} A_i(z) & C_i(z) \\ B_i(z) & D_i(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1} \end{bmatrix}\cdots\begin{bmatrix} 1 & \rho_M z^{-1} \\ \rho_M & z^{-1} \end{bmatrix}\begin{bmatrix} 1 & \rho_{M+1} \\ \rho_{M+1} & 1 \end{bmatrix} \tag{6.6.12}$$

Because there are $M+1-i$ matrix factors that are first-order in z^{-1}, the quantities $A_i(z)$, $B_i(z)$, $C_i(z)$, and $D_i(z)$ will be *polynomials* of order $M+1-i$ in the variable z^{-1}. We may also express (6.6.12) in terms of the transition matrices $F_i(z)$:

$$\begin{bmatrix} A_i(z) & C_i(z) \\ B_i(z) & D_i(z) \end{bmatrix} = z^{-(M+1-i)/2}\nu_i F_i(z)\cdots F_M(z)F_{M+1} \tag{6.6.13}$$

It follows from Eq. (6.6.7) that (6.6.13) will also satisfy similar properties. Indeed, it can be shown easily that:

$$\begin{aligned}\bar{G}_i(z)^T J_3 G_i(z) &= \sigma_i^2 J_3, \qquad \text{where} \quad \sigma_i^2 = \prod_{m=i}^{M+1}(1-\rho_m^2) \\ G_i^R(z) &= J_1 G_i(z) J_1\end{aligned} \tag{6.6.14}$$

where $G_i(z)$ and its reverse $G_i^R(z)$ consisting of the *reversed* polynomials are:

$$G_i(z) = \begin{bmatrix} A_i(z) & C_i(z) \\ B_i(z) & D_i(z) \end{bmatrix}, \quad G_i^R(z) = \begin{bmatrix} A_i^R(z) & C_i^R(z) \\ B_i^R(z) & D_i^R(z) \end{bmatrix} \tag{6.6.15}$$

The reverse of a polynomial is obtained by reversing its coefficients, for example, if $A(z)$ has coefficient vector $\mathbf{a} = [a_0, a_1, a_2, a_3]$, then $A^R(z)$ will have coefficient vector $\mathbf{a}^R = [a_3, a_2, a_1, a_0]$. The reverse of a polynomial can be obtained directly in the z-domain by the property:

$$A^R(z) = z^{-d}A(z^{-1}) = z^{-d}\bar{A}(z)$$

where d is the degree of the polynomial. For example, we have:

$$\begin{aligned}A(z) &= a_0 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} \\ A^R(z) &= a_3 + a_2 z^{-1} + a_1 z^{-2} + a_0 z^{-3} = z^{-3}(a_0 + a_1 z + a_2 z^2 + a_3 z^3) = z^{-3}\bar{A}(z)\end{aligned}$$

Writing the second of Eqs. (6.6.14) explicitly, we have:

$$\begin{bmatrix} A_i^R(z) & C_i^R(z) \\ B_i^R(z) & D_i^R(z) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} A_i(z) & C_i(z) \\ B_i(z) & D_i(z) \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} D_i(z) & B_i(z) \\ C_i(z) & A_i(z) \end{bmatrix}$$

This implies that the polynomials $C_i(z), D_i(z)$ are the reverse of $B_i(z), A_i(z)$, that is, $C_i(z) = B_i^R(z)$, $D_i(z) = A_i^R(z)$. Using this result, the first of Eqs. (6.6.14) implies the following constraint between $A_i(z)$ and $B_i(z)$:

$$\boxed{\bar{A}_i(z)A_i(z) - \bar{B}_i(z)B_i(z) = \sigma_i^2} \tag{6.6.16}$$

Thus, the product of matrices in Eq. (6.6.12) has the form:

$$\begin{bmatrix} A_i(z) & B_i^R(z) \\ B_i(z) & A_i^R(z)\end{bmatrix} = \begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1}\end{bmatrix}\cdots\begin{bmatrix} 1 & \rho_M z^{-1} \\ \rho_M & z^{-1}\end{bmatrix}\begin{bmatrix} 1 & \rho_{M+1} \\ \rho_{M+1} & 1\end{bmatrix} \tag{6.6.17}$$

This definition implies also the recursion:

$$\begin{bmatrix} A_i(z) & B_i^R(z) \\ B_i(z) & A_i^R(z)\end{bmatrix} = \begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1}\end{bmatrix}\begin{bmatrix} A_{i+1}(z) & B_{i+1}^R(z) \\ B_{i+1}(z) & A_{i+1}^R(z)\end{bmatrix} \tag{6.6.18}$$

Therefore, each column will satisfy the same recursion:[†]

$$\begin{bmatrix} A_i(z) \\ B_i(z)\end{bmatrix} = \begin{bmatrix} 1 & \rho_i z^{-1} \\ \rho_i & z^{-1}\end{bmatrix}\begin{bmatrix} A_{i+1}(z) \\ B_{i+1}(z)\end{bmatrix} \quad \text{(forward recursion)} \tag{6.6.19}$$

for $i = M, M-1, \dots, 1$, and initialized by the 0th degree polynomials:

$$\begin{bmatrix} A_{M+1}(z) \\ B_{M+1}(z)\end{bmatrix} = \begin{bmatrix} 1 \\ \rho_{M+1}\end{bmatrix} \tag{6.6.20}$$

Eq. (6.6.11) reads now:

$$\begin{bmatrix} E_{i+} \\ E_{i-}\end{bmatrix} = \frac{z^{(M+1-i)/2}}{\nu_i}\begin{bmatrix} A_i(z) & B_i^R(z) \\ B_i(z) & A_i^R(z)\end{bmatrix}\begin{bmatrix} E'_{M+1,+} \\ E'_{M+1,-}\end{bmatrix} \tag{6.6.21}$$

Setting $i = 1$, we find the relationship between the fields incident on the dielectric structure from the left to those incident from the right:

$$\begin{bmatrix} E_{1+} \\ E_{1-}\end{bmatrix} = \frac{z^{M/2}}{\nu_1}\begin{bmatrix} A_1(z) & B_1^R(z) \\ B_1(z) & A_1^R(z)\end{bmatrix}\begin{bmatrix} E'_{M+1,+} \\ E'_{M+1,-}\end{bmatrix} \tag{6.6.22}$$

where $\nu_1 = \tau_1\tau_2\cdots\tau_{M+1}$. The polynomials $A_1(z)$ and $B_1(z)$ have degree M and are obtained by the recursion (6.6.19). These polynomials incorporate all the multiple reflections and reverberatory effects of the structure.

In referring to the overall transition matrix of the structure, we may drop the subscripts 1 and $M+1$ and write Eq. (6.6.22) in the more convenient form:

$$\begin{bmatrix} E_{+} \\ E_{-}\end{bmatrix} = \frac{z^{M/2}}{\nu}\begin{bmatrix} A(z) & B^R(z) \\ B(z) & A^R(z)\end{bmatrix}\begin{bmatrix} E'_{+} \\ E'_{-}\end{bmatrix} \quad \text{(transfer matrix)} \tag{6.6.23}$$

Fig. 6.6.2 shows the general case of left- and right-incident fields, as well as when the fields are incident only from the left or only from the right.

For both the left- and right-incident cases, the corresponding reflection and transmission responses $\Gamma, \mathcal{T}$ and $\Gamma', \mathcal{T}'$ will satisfy Eq. (6.6.23):

$$\begin{aligned} \begin{bmatrix} 1 \\ \Gamma\end{bmatrix} &= \frac{z^{M/2}}{\nu}\begin{bmatrix} A(z) & B^R(z) \\ B(z) & A^R(z)\end{bmatrix}\begin{bmatrix} \mathcal{T} \\ 0\end{bmatrix} \\ \begin{bmatrix} 0 \\ \mathcal{T}'\end{bmatrix} &= \frac{z^{M/2}}{\nu}\begin{bmatrix} A(z) & B^R(z) \\ B(z) & A^R(z)\end{bmatrix}\begin{bmatrix} \Gamma' \\ 1\end{bmatrix} \end{aligned} \tag{6.6.24}$$

[†] Forward means order-increasing: as the index i decreases, the polynomial order $M+1-i$ increases.

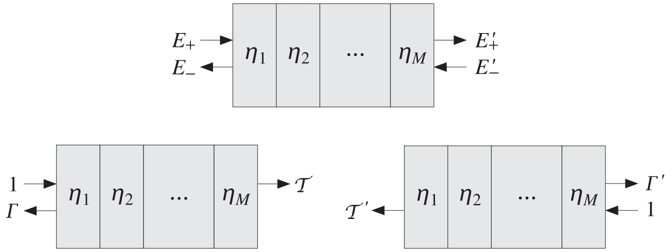


Fig. 6.6.2 Reflection and transmission responses of a multilayer structure.

Solving for $\Gamma, \mathcal{T}$, we find:

$$\boxed{\Gamma(z) = \frac{B(z)}{A(z)}, \quad \mathcal{T}(z) = \frac{\nu z^{-M/2}}{A(z)}} \tag{6.6.25}$$

Similarly, we find for $\Gamma', \mathcal{T}'$:

$$\boxed{\Gamma'(z) = -\frac{B^R(z)}{A(z)}, \quad \mathcal{T}'(z) = \frac{\nu' z^{-M/2}}{A(z)}} \tag{6.6.26}$$

where the constants ν and ν' are the products of the left and right transmission coefficients $\tau_i = 1 + \rho_i$ and $\tau'_i = 1 - \rho_i$, that is,

$$\nu = \prod_{i=1}^{M+1}\tau_i = \prod_{i=1}^{M+1}(1+\rho_i), \quad \nu' = \prod_{i=1}^{M+1}\tau'_i = \prod_{i=1}^{M+1}(1-\rho_i) \tag{6.6.27}$$

In deriving the expression for $\mathcal{T}'$, we used the result (6.6.16), which for $i = 1$ reads:

$$\bar{A}(z)A(z) - \bar{B}(z)B(z) = \sigma^2, \qquad \text{where} \quad \sigma^2 = \prod_{i=1}^{M+1}(1-\rho_i^2) \tag{6.6.28}$$

Because $A^R(z) = z^{-M}\bar{A}(z)$, we can rewrite (6.6.28) in the form:

$$A(z)A^R(z) - B(z)B^R(z) = \sigma^2 z^{-M} \tag{6.6.29}$$

Noting that $\nu\nu' = \sigma^2$ and that

$$\frac{\nu'}{\nu} = \prod_{i=1}^{M+1}\frac{1-\rho_i}{1+\rho_i} = \prod_{i=1}^{M+1}\frac{\eta_{i-1}}{\eta_i} = \frac{\eta_a}{\eta_b},$$

we may replace ν and ν' by the more convenient forms:

$$\nu = \sigma\sqrt{\frac{\eta_b}{\eta_a}}, \quad \nu' = \sigma\sqrt{\frac{\eta_a}{\eta_b}} \tag{6.6.30}$$

Then, the transmission responses $\mathcal{T}$ and $\mathcal{T}'$ can be expressed as:

$$\boxed{\mathcal{T}(z)=\sqrt{\frac{\eta_b}{\eta_a}}\,T(z)\,,\quad \mathcal{T}'(z)=\sqrt{\frac{\eta_a}{\eta_b}}\,T(z)\,,\quad T(z)=\frac{\sigma z^{-M/2}}{A(z)}} \tag{6.6.31}$$

The magnitude squared of $T(z)$ represents the transmittance, that is, the ratio of the transmitted to incident powers, whereas $\mathcal{T}$ is the corresponding ratio of the electric fields. Indeed, assuming $E'_- = 0$, we have $\mathcal{T} = E'_+/E_+$ and find:

$$\frac{\mathcal{P}_{\text{transmitted}}}{\mathcal{P}_{\text{incident}}} = \frac{\dfrac{1}{2\eta_b}\,|E'_+|^2}{\dfrac{1}{2\eta_a}\,|E_+|^2} = \frac{\eta_a}{\eta_b}\,|\mathcal{T}|^2 = |T|^2 \tag{6.6.32}$$

where we used Eq. (6.6.31). Similarly, if the incident fields are from the right, then assuming $E_+ = 0$, the corresponding transmission coefficient will be $\mathcal{T}' = E_-/E'_-$, and we find for the left-going transmittance:

$$\frac{\mathcal{P}'_{\text{transmitted}}}{\mathcal{P}'_{\text{incident}}} = \frac{\dfrac{1}{2\eta_a}\,|E_-|^2}{\dfrac{1}{2\eta_b}\,|E'_-|^2} = \frac{\eta_b}{\eta_a}\,|\mathcal{T}'|^2 = |T|^2 \tag{6.6.33}$$

Eqs. (6.6.32) and (6.6.33) state that the transmittance is the *same* from either side of the structure. This result remains valid even when the slabs are lossy.

The *frequency response* of the structure is obtained by setting $z = e^{j\omega T_s}$. Denoting $A(e^{j\omega T_s})$ simply by $A(\omega)$, we may express Eq. (6.6.28) in the form:

$$|A(\omega)|^2 - |B(\omega)|^2 = \sigma^2 \tag{6.6.34}$$

This implies the following relationship between reflectance and transmittance:

$$\boxed{|\Gamma(\omega)|^2 + |T(\omega)|^2 = 1} \tag{6.6.35}$$

Indeed, dividing Eq. (6.6.34) by $|A(\omega)|^2$ and using Eq. (6.6.31), we have:

$$1 - \left|\frac{B(\omega)}{A(\omega)}\right|^2 = \frac{\sigma^2}{|A(\omega)|^2} = \left|\frac{\sigma e^{-jM\omega T_s/2}}{A(\omega)}\right|^2 \quad\Rightarrow\quad 1 - |\Gamma(\omega)|^2 = |T(\omega)|^2$$

Scattering Matrix

The transfer matrix in Eq. (6.6.23) relates the incident and reflected fields at the left of the structure to those at the right of the structure. Using Eqs. (6.6.25), (6.6.26), and (6.6.29), we may rearrange the transfer matrix (6.6.23) into a *scattering matrix* form that relates the *incoming* fields E_+, E'_- to the *outgoing* fields E_-, E'_+. We have:

$$\begin{bmatrix} E_- \\ E'_+ \end{bmatrix} = \begin{bmatrix} \Gamma(z) & \mathcal{T}'(z) \\ \mathcal{T}(z) & \Gamma'(z) \end{bmatrix} \begin{bmatrix} E_+ \\ E'_- \end{bmatrix} \qquad \text{(scattering matrix)} \tag{6.6.36}$$

The elements of the scattering matrix are referred to as the *S-parameters* and are used widely in the characterization of two-port (and multi-port) networks at microwave frequencies.

We discuss S-parameters in Sec. 13.1. It is a common convention in the literature to normalize the fields to the impedances of the left and right media (the generator and load impedances), as follows:

$$\mathcal{E}_\pm = \frac{1}{\sqrt{\eta_a}}\,E_\pm = \frac{E \pm \eta_a H}{2\sqrt{\eta_a}}\,, \qquad \mathcal{E}'_\pm = \frac{1}{\sqrt{\eta_b}}\,E'_\pm = \frac{E' \pm \eta_b H'}{2\sqrt{\eta_b}} \tag{6.6.37}$$

Such normalized fields are referred to as *power waves* [960]. Using the results of Eq. (6.6.31), the scattering matrix may be written in terms of the normalized fields in the more convenient form:

$$\begin{bmatrix} \mathcal{E}_- \\ \mathcal{E}'_+ \end{bmatrix} = \begin{bmatrix} \Gamma(z) & T(z) \\ T(z) & \Gamma'(z) \end{bmatrix} \begin{bmatrix} \mathcal{E}_+ \\ \mathcal{E}'_- \end{bmatrix} = S(z) \begin{bmatrix} \mathcal{E}_+ \\ \mathcal{E}'_- \end{bmatrix} \tag{6.6.38}$$

so that $S(z)$ is now a symmetric matrix:

$$S(z) = \begin{bmatrix} \Gamma(z) & T(z) \\ T(z) & \Gamma'(z) \end{bmatrix} \qquad \text{(scattering matrix)} \tag{6.6.39}$$

One can verify also that Eqs. (6.6.25), (6.6.26), and (6.6.28) imply the following unitarity properties of $S(z)$:

$$\bar{S}(z)^T S(z) = I\,, \qquad S(\omega)^\dagger S(\omega) = I\,, \qquad \text{(unitarity)} \tag{6.6.40}$$

where I is the $2{\times}2$ identity matrix, $\bar{S}(z) = S(z^{-1})$, and $S(\omega)$ denotes $S(z)$ with $z = e^{j\omega T_s}$, so that $\bar{S}(\omega)^T$ becomes the hermitian conjugate $S(\omega)^\dagger = S(\omega)^{*T}$.

The unitarity condition is equivalent to the power conservation condition that the net incoming power into the (lossless) multilayer structure is equal to the net outgoing reflected power from the structure. Indeed, in terms of the power waves, we have:

$$\begin{aligned}
\mathcal{P}_{\text{out}} &= \frac{1}{2\eta_a}|E_-|^2 + \frac{1}{2\eta_b}|E'_+|^2 = \frac{1}{2}|\mathcal{E}_-|^2 + \frac{1}{2}|\mathcal{E}'_+|^2 \\
&= \frac{1}{2}[\mathcal{E}_-^*, \mathcal{E}_+^{*\prime}] \begin{bmatrix} \mathcal{E}_- \\ \mathcal{E}'_+ \end{bmatrix} = \frac{1}{2}[\mathcal{E}_+^*, \mathcal{E}_-^{*\prime}] S^\dagger S \begin{bmatrix} \mathcal{E}_+ \\ \mathcal{E}'_- \end{bmatrix} = \frac{1}{2}[\mathcal{E}_+^*, \mathcal{E}_-^{*\prime}] I \begin{bmatrix} \mathcal{E}_+ \\ \mathcal{E}'_- \end{bmatrix} \\
&= \frac{1}{2}|\mathcal{E}_+|^2 + \frac{1}{2}|\mathcal{E}'_-|^2 = \frac{1}{2\eta_a}|E_+|^2 + \frac{1}{2\eta_b}|E'_-|^2 = \mathcal{P}_{\text{in}}
\end{aligned}$$

Layer Recursions

Next, we discuss the layer recursions. The reflection responses at the successive interfaces of the structure are given by similar equations to (6.6.25). We have $\Gamma_i(z) = B_i(z)/A_i(z)$ at the ith interface and $\Gamma_{i+1}(z) = B_{i+1}(z)/A_{i+1}(z)$ at the next one. Using Eq. (6.6.19), we find that the responses Γ_i satisfy the following recursion, which is equivalent to Eq. (6.1.3):

$$\boxed{\Gamma_i(z) = \frac{\rho_i + z^{-1}\Gamma_{i+1}(z)}{1 + \rho_i z^{-1}\Gamma_{i+1}(z)}}\,, \qquad i = M, M-1, \dots, 1 \tag{6.6.41}$$

It starts at $\Gamma_{M+1}(z)=\rho_{M+1}$ and ends with $\Gamma(z)=\Gamma_1(z)$. The impedances at the interfaces satisfy Eq. (6.1.5), which takes the specialized form in the case of equal phase thicknesses:

$$\boxed{Z_i(s)=\eta_i\,\frac{Z_{i+1}(s)+\eta_i s}{\eta_i+sZ_{i+1}(s)}}\,,\quad i=M,M-1,\dots,1 \tag{6.6.42}$$

where we defined the variable s via the *bilinear transformation*:

$$s=\frac{1-z^{-1}}{1+z^{-1}} \tag{6.6.43}$$

Note that if $z=e^{2j\delta}$, then $s=j\tan\delta$. It is more convenient to think of the impedances $Z_i(s)$ as functions of the variable s and the reflection responses $\Gamma_i(z)$ as functions of the variable z.

To summarize, given the characteristic impedances $\{\eta_a,\eta_1,\dots,\eta_M,\eta_b\}$, equivalently, the refractive indices $\{n_1,n_1,\dots,n_M\}$ of a multilayered structure, we can compute the corresponding reflection coefficients $\{\rho_1,\rho_2,\dots,\rho_{M+1}\}$ and then carry out the polynomial recursions (6.6.19), eventually arriving at the final Mth order polynomials $A(z)$ and $B(z)$, which define via Eq. (6.6.25) the overall reflection and transmission responses of the structure.

Conversely, given the final polynomials $A_1(z)=A(z)$ and $B_1(z)=B(z)$, we invert the recursion (6.6.19) and "peel off" one layer at a time, until we arrive at the rightmost interface. In the process, we extract the reflection coefficients $\{\rho_1,\rho_2,\dots,\rho_{M+1}\}$, as well as the characteristic impedances and refractive indices of the structure.

This inverse recursion is based on the property that the reflection coefficients appear in the first and last coefficients of the polynomials $B_i(z)$ and $A_i(z)$. Indeed, if we define these coefficients by the expansions:

$$B_i(z)=\sum_{m=0}^{M+1-i}b_i(m)z^{-m}\,,\quad A_i(z)=\sum_{m=0}^{M+1-i}a_i(m)z^{-m}$$

then, it follows from Eq. (6.6.19) that the first coefficients are:

$$b_i(0)=\rho_i\,,\quad a_i(0)=1 \tag{6.6.44}$$

whereas the last coefficients are:

$$b_i(M+1-i)=\rho_{M+1}\,,\quad a_i(M+1-i)=\rho_{M+1}\rho_i \tag{6.6.45}$$

Inverting the transition matrix in Eq. (6.6.19), we obtain the backward recursion:[†]

$$\begin{bmatrix}A_{i+1}(z)\\ B_{i+1}(z)\end{bmatrix}=\frac{1}{1-\rho_i^2}\begin{bmatrix}1 & -\rho_i\\ -\rho_i z & z\end{bmatrix}\begin{bmatrix}A_i(z)\\ B_i(z)\end{bmatrix}\quad\text{(backward recursion)} \tag{6.6.46}$$

[†] Backward means order-decreasing: as the index i increases, the polynomial order $M+1-i$ decreases.

for $i=1,2,\dots,M$, where $\rho_i=b_i(0)$. This recursion starts with the knowledge of $A_1(z)$ and $B_1(z)$. We note that each step of the recursion reduces the order of the polynomials by one, until we reach the 0th order polynomials $A_{M+1}(z)=1$ and $B_{M+1}(z)=\rho_{M+1}$.

The reverse recursions can also be applied directly to the reflection responses $\Gamma_i(z)$ and wave impedances $Z_i(s)$. It follows from Eq. (6.6.41) that the reflection coefficient ρ_i can be extracted from $\Gamma_i(z)$ if we set $z=\infty$, that is, $\rho_i=\Gamma_i(\infty)$. Then, solving Eq. (6.1.3) for $\Gamma_{i+1}(z)$, we obtain:

$$\boxed{\Gamma_{i+1}(z)=z\,\frac{\Gamma_i(z)-\rho_i}{1-\rho_i\Gamma_i(z)}}\,,\quad i=1,2,\dots,M \tag{6.6.47}$$

Similarly, it follows from Eq. (6.6.42) that the characteristic impedance η_i can be extracted from $Z_i(s)$ by setting $s=1$, which is equivalent to $z=\infty$ under the transformation (6.6.43). Thus, $\eta_i=Z_i(1)$ and the inverse of (6.6.42) becomes:

$$\boxed{Z_{i+1}(s)=\eta_i\,\frac{Z_i(s)-s\eta_i}{\eta_i-sZ_i(s)}}\,,\quad i=1,2,\dots,M \tag{6.6.48}$$

The necessary and sufficient condition that the extracted reflection coefficients ρ_i and the media impedances η_i are realizable, that is, $|\rho_i|<1$ or $\eta_i>0$, is that the starting polynomial $A(z)$ be a *minimum-phase polynomial* in z^{-1}, that is, it must have all its zeros inside the unit circle on the z-plane. This condition is in turn equivalent to the requirement that the transmission and reflection responses $T(z)$ and $\Gamma(z)$ be *stable and causal* transfer functions.

The order-increasing and order-decreasing recursions Eqs. (6.6.19) and (6.6.46) can also be expressed in terms of the vectors of coefficients of the polynomials $A_i(z)$ and $B_i(z)$. Defining the column vectors:

$$\mathbf{a}_i=\begin{bmatrix}a_i(0)\\ a_i(1)\\ \vdots\\ a_i(M+1-i)\end{bmatrix},\quad \mathbf{b}_i=\begin{bmatrix}b_i(0)\\ b_i(1)\\ \vdots\\ b_i(M+1-i)\end{bmatrix}$$

we obtain for Eq. (6.6.19), with $i=M,M-1,\dots,1$:

$$\begin{aligned}\mathbf{a}_i&=\begin{bmatrix}\mathbf{a}_{i+1}\\ 0\end{bmatrix}+\rho_i\begin{bmatrix}0\\ \mathbf{b}_{i+1}\end{bmatrix}\\ \mathbf{b}_i&=\rho_i\begin{bmatrix}\mathbf{a}_{i+1}\\ 0\end{bmatrix}+\begin{bmatrix}0\\ \mathbf{b}_{i+1}\end{bmatrix}\end{aligned}\quad\text{(forward recursion)} \tag{6.6.49}$$

and initialized at $\mathbf{a}_{M+1}=[1]$ and $\mathbf{b}_{M+1}=[\rho_{M+1}]$. Similarly, the backward recursions (6.6.46) are initialized at the Mth order polynomials $\mathbf{a}_1=\mathbf{a}$ and $\mathbf{b}_1=\mathbf{b}$. For $i=1,2,\dots,M$ and $\rho_i=b_i(0)$, we have:

$$\begin{aligned}\begin{bmatrix}\mathbf{a}_{i+1}\\ 0\end{bmatrix}&=\frac{\mathbf{a}_i-\rho_i\mathbf{b}_i}{1-\rho_i^2}\\ \begin{bmatrix}0\\ \mathbf{b}_{i+1}\end{bmatrix}&=\frac{-\rho_i\mathbf{a}_i+\mathbf{b}_i}{1-\rho_i^2}\end{aligned}\quad\text{(backward recursion)} \tag{6.6.50}$$

Example 6.6.1: Determine the number of layers M, the reflection coefficients at the $M+1$ interfaces, and the refractive indices of the $M+2$ media for a multilayer structure whose overall reflection response is given by:

$$\Gamma(z)=\frac{B(z)}{A(z)}=\frac{-0.1-0.188z^{-1}-0.35z^{-2}+0.5z^{-3}}{1-0.1z^{-1}-0.064z^{-2}-0.05z^{-3}}$$

Solution: From the degree of the polynomials, the number of layers is $M=3$. The starting polynomials in the backward recursion (6.6.50) are:

$$\mathbf{a}_1=\mathbf{a}=\begin{bmatrix}1.000\\-0.100\\-0.064\\-0.050\end{bmatrix},\quad \mathbf{b}_1=\mathbf{b}=\begin{bmatrix}-0.100\\-0.188\\-0.350\\0.500\end{bmatrix}$$

From the first and last coefficients of $\mathbf{b}_1$, we find $\rho_1=-0.1$ and $\rho_4=0.5$. Setting $i=1$, the first step of the recursion gives:

$$\begin{bmatrix}\mathbf{a}_2\\0\end{bmatrix}=\frac{\mathbf{a}_1-\rho_1\mathbf{b}_1}{1-\rho_1^2}=\begin{bmatrix}1.000\\-0.120\\-0.100\\0.000\end{bmatrix},\quad \begin{bmatrix}0\\\mathbf{b}_2\end{bmatrix}=\frac{-\rho_1\mathbf{a}_1+\mathbf{b}_1}{1-\rho_1^2}=\begin{bmatrix}0.000\\-0.200\\-0.360\\0.500\end{bmatrix}$$

Thus,

$$\mathbf{a}_2=\begin{bmatrix}1.000\\-0.120\\-0.100\end{bmatrix},\quad \mathbf{b}_2=\begin{bmatrix}-0.200\\-0.360\\0.500\end{bmatrix}$$

The first coefficient of $\mathbf{b}_2$ is $\rho_2=-0.2$ and the next step of the recursion gives:

$$\begin{bmatrix}\mathbf{a}_3\\0\end{bmatrix}=\frac{\mathbf{a}_2-\rho_2\mathbf{b}_2}{1-\rho_2^2}=\begin{bmatrix}1.0\\-0.2\\0.0\end{bmatrix},\quad \begin{bmatrix}0\\\mathbf{b}_3\end{bmatrix}=\frac{-\rho_2\mathbf{a}_2+\mathbf{b}_2}{1-\rho_2^2}=\begin{bmatrix}0.0\\-0.4\\0.5\end{bmatrix}$$

Thus,

$$\mathbf{a}_3=\begin{bmatrix}1.0\\-0.2\end{bmatrix},\quad \mathbf{b}_3=\begin{bmatrix}-0.4\\0.5\end{bmatrix}\quad\Rightarrow\quad \rho_3=-0.4$$

The last step of the recursion for $i=3$ is not necessary because we have already determined $\rho_4=0.5$. Thus, the four reflection coefficients are:

$$[\rho_1,\rho_2,\rho_3,\rho_4]=[-0.1,-0.2,-0.4,0.5]$$

The corresponding refractive indices can be obtained by solving Eq. (6.1.1), that is, $n_i=n_{i-1}(1-\rho_i)/(1+\rho_i)$. Starting with $i=1$ and $n_0=n_a=1$, we obtain:

$$[n_a,n_1,n_2,n_3,n_b]=[1,1.22,1.83,4.28,1.43]$$

The same results can be obtained by working with the polynomial version of the recursion, Eq. (6.6.46). □

Example 6.6.2: Consider the quarter-quarter antireflection coating shown in Fig. 6.2.2 with refractive indices $[n_a,n_1,n_2,n_b]=[1,1.38,1.63,1.50]$. Determine the reflection coefficients at the three interfaces and the overall reflection response $\Gamma(z)$ of the structure.

Solution: In this problem we carry out the forward layer recursion starting from the rightmost layer. The reflection coefficients computed from Eq. (6.1.1) are:

$$[\rho_1,\rho_2,\rho_3]=[-0.1597,-0.0831,0.0415]$$

Starting the forward recursion with $\mathbf{a}_3=[1]$ and $\mathbf{b}_3=[\rho_3]=[0.0415]$, we build the first order polynomials:

$$\mathbf{a}_2=\begin{bmatrix}\mathbf{a}_3\\0\end{bmatrix}+\rho_2\begin{bmatrix}0\\\mathbf{b}_3\end{bmatrix}=\begin{bmatrix}1.0000\\0.0000\end{bmatrix}+(-0.0831)\begin{bmatrix}0.0000\\0.0415\end{bmatrix}=\begin{bmatrix}1.0000\\-0.0034\end{bmatrix}$$

$$\mathbf{b}_2=\rho_2\begin{bmatrix}\mathbf{a}_3\\0\end{bmatrix}+\begin{bmatrix}0\\\mathbf{b}_3\end{bmatrix}=(-0.0831)\begin{bmatrix}1.0000\\0.0000\end{bmatrix}+\begin{bmatrix}0.0000\\0.0415\end{bmatrix}=\begin{bmatrix}-0.0831\\0.0415\end{bmatrix}$$

Then, we build the 2nd order polynomials at the first interface:

$$\mathbf{a}_1=\begin{bmatrix}\mathbf{a}_2\\0\end{bmatrix}+\rho_1\begin{bmatrix}0\\\mathbf{b}_2\end{bmatrix}=\begin{bmatrix}1.0000\\0.0098\\-0.0066\end{bmatrix},\quad \mathbf{b}_1=\rho_1\begin{bmatrix}\mathbf{a}_2\\0\end{bmatrix}+\begin{bmatrix}0\\\mathbf{b}_2\end{bmatrix}=\begin{bmatrix}-0.1597\\-0.0825\\0.0415\end{bmatrix}$$

Thus, the overall reflection response is:

$$\Gamma(z)=\Gamma_1(z)=\frac{B_1(z)}{A_1(z)}=\frac{-0.1597-0.0825z^{-1}+0.0415z^{-2}}{1+0.0098z^{-1}-0.0066z^{-2}}$$

Applying the reverse recursion on this reflection response would generate the same reflection coefficients ρ_1,ρ_2,ρ_3. □

Example 6.6.3: Determine the overall reflection response of the quarter-half-quarter coating of Fig. 6.2.2 by thinking of the half-wavelength layer as two quarter-wavelength layers of the same refractive index.

Solution: There are $M=4$ quarter-wave layers with refractive indices:

$$[n_a,n_1,n_2,n_3,n_4,n_b]=[1,1.38,2.20,2.20,1.63,1.50]$$

The corresponding reflection coefficients are:

$$[\rho_1,\rho_2,\rho_3,\rho_4,\rho_5]=[-0.1597,-0.2291,0,0.1488,0.0415]$$

where the reflection coefficient at the imaginary interface separating the two halves of the half-wave layer is zero. Starting the forward recursion with $\mathbf{a}_5=[1]$, $\mathbf{b}_5=[\rho_5]=[0.0415]$, we compute the higher-order polynomials:

$$\mathbf{a}_4=\begin{bmatrix}\mathbf{a}_5\\0\end{bmatrix}+\rho_4\begin{bmatrix}0\\\mathbf{b}_5\end{bmatrix}=\begin{bmatrix}1.0000\\0.0062\end{bmatrix},\quad \mathbf{b}_4=\rho_4\begin{bmatrix}\mathbf{a}_5\\0\end{bmatrix}+\begin{bmatrix}0\\\mathbf{b}_5\end{bmatrix}=\begin{bmatrix}0.1488\\0.0415\end{bmatrix}$$

$$\mathbf{a}_3 = \begin{bmatrix} \mathbf{a}_4 \\ 0 \end{bmatrix} + \rho_3 \begin{bmatrix} 0 \\ \mathbf{b}_4 \end{bmatrix} = \begin{bmatrix} 1.0000 \\ 0.0062 \\ 0.0000 \end{bmatrix}, \quad \mathbf{b}_3 = \rho_3 \begin{bmatrix} \mathbf{a}_4 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{b}_4 \end{bmatrix} = \begin{bmatrix} 0.0000 \\ 0.1488 \\ 0.0415 \end{bmatrix}$$

$$\mathbf{a}_2 = \begin{bmatrix} \mathbf{a}_3 \\ 0 \end{bmatrix} + \rho_2 \begin{bmatrix} 0 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 1.0000 \\ 0.0062 \\ -0.0341 \\ -0.0095 \end{bmatrix}, \quad \mathbf{b}_2 = \rho_2 \begin{bmatrix} \mathbf{a}_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} -0.2291 \\ -0.0014 \\ 0.1488 \\ 0.0415 \end{bmatrix}$$

$$\mathbf{a}_1 = \begin{bmatrix} \mathbf{a}_2 \\ 0 \end{bmatrix} + \rho_1 \begin{bmatrix} 0 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 1.0000 \\ 0.0428 \\ -0.0339 \\ -0.0333 \\ -0.0066 \end{bmatrix}, \quad \mathbf{b}_1 = \rho_1 \begin{bmatrix} \mathbf{a}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} -0.1597 \\ -0.2300 \\ 0.0040 \\ 0.1503 \\ 0.0415 \end{bmatrix}$$

Thus, the reflection response will be:

$$\Gamma(z) = \frac{B_1(z)}{A_1(z)} = \frac{-0.1597 - 0.2300z^{-1} + 0.0040z^{-2} + 0.1502z^{-3} + 0.0415z^{-4}}{1 + 0.0428z^{-1} - 0.0339z^{-2} - 0.0333z^{-3} - 0.0066z^{-4}}$$

We note that because $\rho_3 = 0$, the polynomials $A_3(z)$ and $A_4(z)$ are the same and $B_3(z)$ is simply the delayed version of $B_4(z)$, that is, $B_3(z) = z^{-1}B_4(z)$. □

Example 6.6.4: Determine the reflection polynomials for the cases $M = 1$, $M = 2$, and $M = 3$ with reflection coefficients $\{\rho_1, \rho_2\}$, $\{\rho_1, \rho_2, \rho_3\}$, and $\{\rho_1, \rho_2, \rho_3, \rho_4\}$, respectively.

Solution: For $M = 1$, we have $A_2(z) = 1$ and $B_2(z) = \rho_2$. Then, Eq. (6.6.19) gives:

$$\begin{bmatrix} A_1(z) \\ B_1(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 z^{-1} \\ \rho_1 & z^{-1} \end{bmatrix} \begin{bmatrix} A_2(z) \\ B_2(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 z^{-1} \\ \rho_1 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 + \rho_1\rho_2 z^{-1} \\ \rho_1 + \rho_2 z^{-1} \end{bmatrix}$$

For $M = 2$, we start with $A_3(z) = 1$ and $B_3(z) = \rho_3$. The first step of the recursion gives:

$$\begin{bmatrix} A_2(z) \\ B_2(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_2 z^{-1} \\ \rho_2 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 1 + \rho_2\rho_3 z^{-1} \\ \rho_2 + \rho_3 z^{-1} \end{bmatrix}$$

and the second step:

$$\begin{bmatrix} A_1(z) \\ B_1(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 z^{-1} \\ \rho_1 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 + \rho_2\rho_3 z^{-1} \\ \rho_2 + \rho_3 z^{-1} \end{bmatrix} = \begin{bmatrix} 1 + \rho_2(\rho_1 + \rho_3)z^{-1} + \rho_1\rho_3 z^{-2} \\ \rho_1 + \rho_2(1 + \rho_1\rho_3)z^{-1} + \rho_3 z^{-2} \end{bmatrix}$$

For $M = 3$, we have $A_4(z) = 1$ and $B_4(z) = \rho_4$. The first and second steps give:

$$\begin{bmatrix} A_3(z) \\ B_3(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_3 z^{-1} \\ \rho_3 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ \rho_4 \end{bmatrix} = \begin{bmatrix} 1 + \rho_3\rho_4 z^{-1} \\ \rho_3 + \rho_4 z^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A_2(z) \\ B_2(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_2 z^{-1} \\ \rho_2 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 + \rho_3\rho_4 z^{-1} \\ \rho_3 + \rho_4 z^{-1} \end{bmatrix} = \begin{bmatrix} 1 + \rho_3(\rho_2 + \rho_4)z^{-1} + \rho_2\rho_4 z^{-2} \\ \rho_2 + \rho_3(1 + \rho_2\rho_4)z^{-1} + \rho_4 z^{-2} \end{bmatrix}$$

Then, the final step gives:

$$\begin{bmatrix} A_1(z) \\ B_1(z) \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 z^{-1} \\ \rho_1 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 + \rho_3(\rho_2 + \rho_4)z^{-1} + \rho_2\rho_4 z^{-2} \\ \rho_2 + \rho_3(1 + \rho_2\rho_4)z^{-1} + \rho_4 z^{-2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 + (\rho_1\rho_2 + \rho_2\rho_3 + \rho_3\rho_4)z^{-1} + (\rho_1\rho_3 + \rho_2\rho_4 + \rho_1\rho_2\rho_3\rho_4)z^{-2} + \rho_1\rho_4 z^{-3} \\ \rho_1 + (\rho_2 + \rho_1\rho_2\rho_3 + \rho_1\rho_3\rho_4)z^{-1} + (\rho_3 + \rho_1\rho_2\rho_4 + \rho_2\rho_3\rho_4)z^{-2} + \rho_4 z^{-3} \end{bmatrix}$$

As expected, in all cases the first and last coefficients of $A_i(z)$ are 1 and $\rho_i\rho_{M+1}$ and those of $B_i(z)$ are ρ_i and ρ_{M+1}.

An approximation that is often made in practice is to assume that the ρ_is are small and ignore all the terms that involve two or more factors of ρ_i. In this approximation, we have for the polynomials and the reflection response $\Gamma(z) = B_1(z)/A_1(z)$, for the $M = 3$ case:

$$\begin{aligned} A_1(z) &= 1 \\ B_1(z) &= \rho_1 + \rho_2 z^{-1} + \rho_3 z^{-2} + \rho_4 z^{-3} \end{aligned} \quad \Rightarrow \quad \Gamma(z) = \rho_1 + \rho_2 z^{-1} + \rho_3 z^{-2} + \rho_4 z^{-3}$$

This is equivalent to *ignoring all multiple reflections* within each layer and considering only a *single* reflection at each interface. Indeed, the term $\rho_2 z^{-1}$ represents the wave reflected at interface-2 and arriving back at interface-1 with a roundtrip delay of z^{-1}. Similarly, $\rho_3 z^{-2}$ represents the reflection at interface-3 and has a delay of z^{-2} because the wave must make a roundtrip of two layers to come back to interface-1, and $\rho_4 z^{-3}$ has three roundtrip delays because the wave must traverse three layers. □

The two MATLAB functions `frwrec` and `bkwrec` implement the forward and backward recursions (6.6.49) and (6.6.50), respectively. They have usage:

```
[A,B] = frwrec(r);              % forward recursion - from r to A,B
[r,A,B] = bkwrec(a,b);          % backward recursion - from a,b to r
```

The input r of `frwrec` represents the vector of the $M + 1$ reflection coefficients and A, B are the $(M + 1)\times(M + 1)$ matrices whose columns are the polynomials $\mathbf{a}_i$ and $\mathbf{b}_i$ (padded with zeros at the end to make them of length $M + 1$.) The inputs a, b of `bkwrec` are the final order-M polynomials $\mathbf{a}, \mathbf{b}$ and the outputs r, A, B have the same meaning as in `frwrec`. We note that the first row of B contains the reflection coefficients r.

The auxiliary functions `r2n` and `n2r` allow one to pass from the reflection coefficient vector r to the refractive index vector n, and conversely. They have usage:

```
n = r2n(r);             % reflection coefficients to refractive indices
r = n2r(n);             % refractive indices to reflection coefficients
```

As an illustration, the MATLAB code:

```
a = [1, -0.1, -0.064, -0.05];
b = [-0.1, -0.188, -0.35, 0.5];
[r,A,B] = bkwrec(a,b);
n = r2n(r);
r = n2r(n);
```

will generate the output of Example 6.6.1:

```
r =
   -0.1000   -0.2000   -0.4000    0.5000
A =
    1.0000    1.0000    1.0000    1.0000
   -0.1000   -0.1200   -0.2000         0
   -0.0640   -0.1000         0         0
   -0.0500         0         0         0
B =
   -0.1000   -0.2000   -0.4000    0.5000
   -0.1880   -0.3600    0.5000         0
   -0.3500    0.5000         0         0
    0.5000         0         0         0
n =
    1.0000    1.2222    1.8333    4.2778    1.4259
r =
   -0.1000   -0.2000   -0.4000    0.5000
```

Conversely, if the above r is the input to `frwrec`, the returned matrices A, B will be identical to the above. The function `r2n` solves Eq. (6.1.1) for n_i and always assumes that the refractive index of the leftmost medium is unity. Once the n_i are known, the function `multidiel` may be used to compute the reflection response at any set of frequencies or wavelengths.

6.7 Applications of Layered Structures

In addition to their application in dielectric thin-film and radome design, layered structures and the corresponding forward and backward layer recursions have a number of applications in other wave propagation problems, such as the design of broadband terminations of transmission lines, the analysis and synthesis of speech, geophysical signal processing for oil exploration, the probing of tissue by ultrasound, and the design of acoustic reflectors for noise control.

It is remarkable also that the same forward and backward recursions (6.6.49) and (6.6.50) are identical (up to reindexing) to the forward and backward Levinson recursions of *linear prediction* [792], with the layer structures being mathematically equivalent to the analysis and synthesis lattice filters. This connection is perhaps the reason behind the great success of linear prediction methods in speech and geophysical signal processing.

Moreover, the forward and backward layer recursions in their reflection forms, Eqs. (6.6.41) and (6.6.47), and impedance forms, Eqs. (6.6.42) and (6.6.48), are the essential mathematical tools for Schur's characterization of *lossless bounded real* functions in the z-plane and Richard's characterization of *positive real* functions in the s-plane and have been applied to network synthesis and to the development of transfer function stability tests, such as the Schur-Cohn test [812–826].

In all wave problems there are always *two* associated propagating field quantities playing the roles of the electric and magnetic fields. For forward-moving waves the ratio of the two field quantities is constant and equal to the *characteristic impedance* of the particular propagation medium for the particular type of wave.

For example, for transmission lines the two field quantities are the voltage and current along the line, for sound waves they are the pressure and particle volume velocity, and for seismic waves, the stress and particle displacement.

A transmission line connected to a multisegment impedance transformer and a load is shown in Fig. 6.7.1. The characteristic impedances of the main line and the segments are Z_a and $Z_1, \dots, Z_M$, and the impedance of the load, Z_b. Here, the impedances $\{Z_a, Z_1, \dots, Z_M, Z_b\}$, play the same role as $\{\eta_a, \eta_1, \dots, \eta_M, \eta_b\}$ in the dielectric stack case.

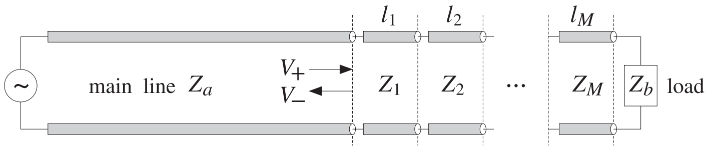


Fig. 6.7.1 Multisegment broadband termination of a transmission line.

The segment characteristic impedances Z_i and lengths l_i can be adjusted to obtain an overall reflection response that is reflectionless over a wideband of frequencies [781–791]. This design method is presented in Sec. 6.8.

In speech processing, the vocal tract is modeled as an acoustic tube of varying cross-sectional area. It can be approximated by the piece-wise constant area approximation shown in Fig. 6.7.2. Typically, ten segments will suffice.

The acoustic impedance of a sound wave varies inversely with the tube area, $Z = \rho c/A$, where ρ, c, and A are the air density, speed of sound, and tube area, respectively. Therefore, as the sound wave propagates from the glottis to the lips, it will suffer reflections every time it encounters an interface, that is, whenever it enters a tube segment of different diameter.

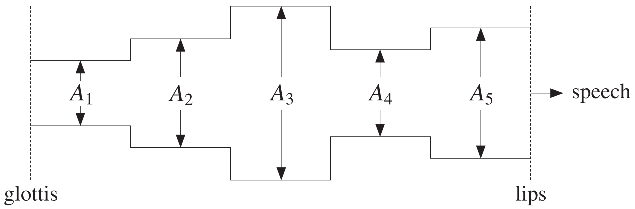


Fig. 6.7.2 Multisegment acoustic tube model of vocal tract.

Multiple reflections will be set up within each segment and the tube will reverberate in a complicated manner depending on the number of segments and their diameters. By measuring the speech wave that eventually comes out of the lips (the transmission response,) it is possible to remove, or deconvolve, the reverberatory effects of the tube and, in the process, extract the tube parameters, such as the areas of the segments, or equivalently, the reflection coefficients at the interfaces.

During speech, the configuration of the vocal tract changes continuously, but it does so at mechanical speeds. For short periods of time (typically, of the order of 20–30 msec,) it may be considered to maintain a fixed configuration. From each such short segment of speech, a set of configuration parameters, such as reflection coefficients, is extracted. Conversely, the extracted parameters may be used to re-synthesize the speech segment.

Such linear prediction based acoustic tube models of speech production are routinely used in the analysis and synthesis of speech, speech recognition, speaker identification, and speech coding for efficient data transmission, such as in wireless phones.

The seismic problem in geophysical signal processing is somewhat different. Here, it is not the transmitted wave that is experimentally available, but rather the overall reflected wave. Fig. 6.7.3 shows the typical case.

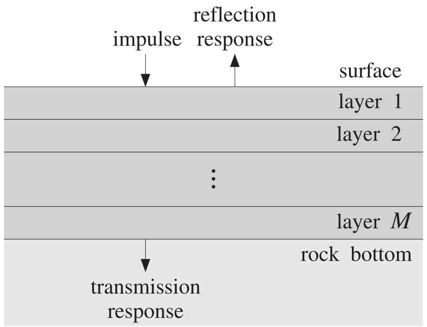


Fig. 6.7.3 Seismic probing of earth's multilayer structure.

An impulsive input to the earth, such as an explosion near the surface, will set up seismic elastic waves propagating downwards. As the various earth layers are encountered, reflections will take place. Eventually, each layer will be reverberating and an overall reflected wave will be measured at the surface. With the help of the backward recursions, the parameters of the layered structure (reflection coefficients and impedances) are extracted and evaluated to determine the presence of a layer that contains an oil deposit.

The application of the backward recursions has been termed *dynamic predictive deconvolution* in the geophysical context [799–811]. An interesting historical account of the early development of this method by Robinson and its application to oil exploration and its connection to linear prediction is given in Ref. [805]. The connection to the conventional inverse scattering methods based on the *Gelfand-Levitan-Marchenko* approach is discussed in [806–811].

Fiber Bragg gratings (FBG), obtained by periodically modulating the refractive index of the core (or the cladding) of a finite portion of a fiber, behave very similarly to dielectric mirrors and exhibit high reflectance bands [747–767]. The periodic modulation is achieved by exposing that portion of the fiber to intense ultraviolet radiation whose intensity has the required periodicity. The periodicity shown in Fig. 6.7.4 can have arbitrary shape—not only alternating high/low refractive index layers as suggested by the figure. We discuss FBGs further in Sec. 11.4.

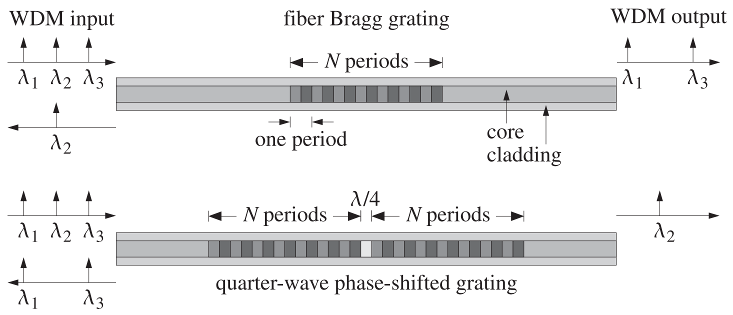


Fig. 6.7.4 Fiber Bragg gratings acting as bandstop or bandpass filters.

Quarter-wave phase-shifted fiber Bragg gratings act as narrow-band transmission filters and can be used as demultiplexing filters in WDM and dense WDM (DWDM) communications systems. Assuming as in Fig. 6.7.4 that the inputs to the FBGs consist of several multiplexed wavelengths, $\lambda_1, \lambda_2, \lambda_3, \dots$, and that the FBGs are tuned to wavelength λ_2, then the ordinary FBG will act as an almost perfect reflector of λ_2. If its reflecting band is narrow, then the other wavelengths will transmit through. Similarly, the phase-shifted FBG will act as a narrow-band transmission filter allowing λ_2 through and reflecting the other wavelengths if they lie within its reflecting band.

A typical DWDM system may carry 40 wavelengths at 10 gigabits per second (Gbps) per wavelength, thus achieving a 400 Gbps bandwidth. In the near future, DWDM systems will be capable of carrying hundreds of wavelengths at 40 Gbps per wavelength, achieving terabit per second rates [767].

6.8 Chebyshev Design of Reflectionless Multilayers

In this section, we discuss the design of broadband reflectionless multilayer structures of the type shown in Fig. 6.6.1 , or equivalently, broadband terminations of transmission lines as shown in Fig. 6.7.1, using Collin's method based on Chebyshev polynomials [781–791,617,636].

As depicted in Fig. 6.8.1, the desired specifications are: (a) the operating center frequency f_0 of the band, (b) the bandwidth Δf, and (c) the desired amount of attenuation A (in dB) within the desired band, measured with respect to the reflectance value at dc.

Because the optical thickness of the layers is $\delta = \omega T_s/2 = (\pi/2)(f/f_0)$ and vanishes at dc, the reflection response at $f = 0$ should be set equal to its unmatched value, that is, to the value when there are no layers:

$$|\Gamma(0)|^2 = \rho_0^2 = \left(\frac{\eta_b - \eta_a}{\eta_a + \eta_b}\right)^2 = \left(\frac{n_a - n_b}{n_a + n_b}\right)^2 \tag{6.8.1}$$

Collin's design method [781] assumes $|\Gamma(f)|^2$ has the analytical form:

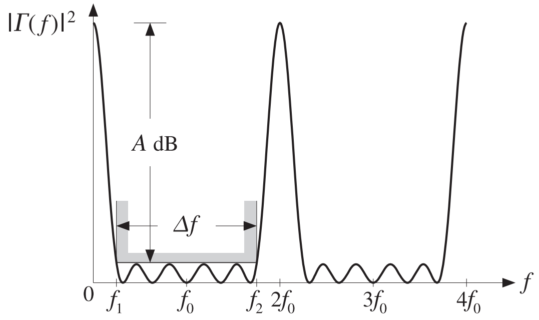


Fig. 6.8.1 Reflectance specifications for Chebyshev design.

$$\boxed{|\Gamma(f)|^2 = \frac{e_1^2 T_M^2(x)}{1 + e_1^2 T_M^2(x)}} \qquad x = x_0\cos\delta = x_0\cos\big(\frac{\pi f}{2f_0}\big) \tag{6.8.2}$$

where $T_M(x) = \cos\big(M\,\mathrm{acos}(x)\big)$ is the Chebyshev polynomial (of the first kind) of order M. The parameters M, e_1, x_0 are fixed by imposing the desired specifications shown in Fig. 6.8.1.

Once these parameters are known, the order-M polynomials $A(z), B(z)$ are determined by spectral factorization, so that $|\Gamma(f)|^2 = |B(f)|^2/|A(f)|^2$. The backward layer recursions, then, allow the determination of the reflection coefficients at the layer interfaces, and the corresponding refractive indices. Setting $f = 0$, or $\delta = 0$, or $\cos\delta = 1$, or $x = x_0$, we obtain the design equation:

$$|\Gamma(0)|^2 = \frac{e_1^2 T_M^2(x_0)}{1 + e_1^2 T_M^2(x_0)} = \frac{e_0^2}{1 + e_0^2} = \rho_0^2 \tag{6.8.3}$$

where we defined $e_0 = e_1 T_M(x_0)$. Solving for e_0, we obtain:

$$\boxed{e_0^2 = \frac{\rho_0^2}{1 - \rho_0^2} = \frac{(n_a - n_b)^2}{4n_a n_b}} \tag{6.8.4}$$

Chebyshev polynomials $T_M(x)$ are reviewed in more detail in Sec. 20.9 that discusses antenna array design using the Dolph-Chebyshev window. The two key properties of these polynomials are that they have *equiripple* behavior within the interval $-1 \le x \le 1$ and grow like x^M for $|x| > 1$; see for example, Fig. 20.9.1.

By adjusting the value of the scale parameter x_0, we can arrange the entire equiripple domain, $-1 \le x \le 1$, of $T_M(x)$ to be mapped onto the desired reflectionless band $[f_1, f_2]$, where f_1, f_2 are the left and right bandedge frequencies about f_0, as shown in Fig. 6.8.1. Thus, we demand the conditions:

$$x_0\cos\big(\frac{\pi f_2}{2f_0}\big) = -1, \qquad x_0\cos\big(\frac{\pi f_1}{2f_0}\big) = 1$$

These can be solved to give:

$$\begin{aligned}
\frac{\pi f_2}{2f_0} &= \mathrm{acos}\big(-\frac{1}{x_0}\big) = \frac{\pi}{2} + \mathrm{asin}\big(\frac{1}{x_0}\big) \\
\frac{\pi f_1}{2f_0} &= \mathrm{acos}\big(\frac{1}{x_0}\big) = \frac{\pi}{2} - \mathrm{asin}\big(\frac{1}{x_0}\big)
\end{aligned} \tag{6.8.5}$$

Subtracting, we obtain the bandwidth $\Delta f = f_2 - f_1$:

$$\frac{\pi}{2}\frac{\Delta f}{f_0} = 2\,\mathrm{asin}\Big(\frac{1}{x_0}\Big) \tag{6.8.6}$$

We can now solve for the scale parameter x_0 in terms of the bandwidth:

$$\boxed{x_0 = \frac{1}{\sin\Big(\dfrac{\pi}{4}\dfrac{\Delta f}{f_0}\Big)}} \tag{6.8.7}$$

It is evident from Fig. 6.8.1 that the maximum value of the bandwidth that one can demand is $\Delta f_{\max} = 2f_0$. Going back to Eq. (6.8.5) and using (6.8.6), we see that f_1 and f_2 lie symmetrically about f_0, such that $f_1 = f_0 - \Delta f/2$ and $f_2 = f_0 + \Delta f/2$.

Next, we impose the attenuation condition. Because of the equiripple behavior over the Δf band, it is enough to impose the condition at the edges of the band, that is, we demand that when $f = f_1$, or $x = 1$, the reflectance is down by A dB as compared to its value at dc:

$$|\Gamma(f_1)|^2 = |\Gamma(0)|^2\,10^{-A/10} \quad\Rightarrow\quad \frac{e_1^2 T_M^2(1)}{1 + e_1^2 T_M^2(1)} = \frac{e_0^2}{1 + e_0^2}\,10^{-A/10}$$

But, $T_M(1) = 1$. Therefore, we obtain an equation for e_1^2:

$$\frac{e_1^2}{1 + e_1^2} = \frac{e_0^2}{1 + e_0^2}\,10^{-A/10} \tag{6.8.8}$$

Noting that $e_0 = e_1 T_M(x_0)$, we solve Eq. (6.8.8) for the ratio $T_M(x_0) = e_0/e_1$:

$$T_M(x_0) = \cosh\big(M\,\mathrm{acosh}(x_0)\big) = \sqrt{(1 + e_0^2)\,10^{A/10} - e_0^2} \tag{6.8.9}$$

Alternatively, we can express A in terms of $T_M(x_0)$:

$$A = 10\log_{10}\left(\frac{T_M^2(x_0) + e_0^2}{1 + e_0^2}\right) \tag{6.8.10}$$

where we used the definition $T_M(x_0) = \cosh\big(M\,\mathrm{acosh}(x_0)\big)$ because $x_0 > 1$. Solving (6.8.9) for M in terms of A, we obtain:

$$\boxed{M = \mathrm{ceil}(M_{\mathrm{exact}})} \tag{6.8.11}$$

where

$$M_{\mathrm{exact}} = \frac{\mathrm{acosh}\Big(\sqrt{(1 + e_0^2)\,10^{A/10} - e_0^2}\Big)}{\mathrm{acosh}(x_0)} \tag{6.8.12}$$

Because M_{exact} is rounded up to the next integer, the attenuation will be somewhat larger than required. In summary, we calculate e_0, x_0, M from Eqs. (6.8.4), (6.8.7), and (6.8.11). Finally, e_1 is calculated from:

$$\boxed{e_1 = \frac{e_0}{T_M(x_0)} = \frac{e_0}{\cosh\big(M\,\text{acosh}(x_0)\big)}} \tag{6.8.13}$$

Next, we construct the polynomials $A(z)$ and $B(z)$. It follows from Eqs. (6.6.25) and (6.6.34) that the reflectance and transmittance are:

$$|\Gamma(f)|^2 = \frac{|B(f)|^2}{|A(f)|^2}\,, \quad |T(f)|^2 = 1 - |\Gamma(f)|^2 = \frac{\sigma^2}{|A(f)|^2}\,,$$

Comparing these with Eq. (6.8.2), we obtain:

$$\begin{aligned} |A(f)|^2 &= \sigma^2\big[1 + e_1^2 T_M^2(x_0\cos\delta)\big] \\ |B(f)|^2 &= \sigma^2 e_1^2 T_M^2(x_0\cos\delta) \end{aligned} \tag{6.8.14}$$

The polynomial $A(z)$ is found by requiring that it be a minimum-phase polynomial, that is, with all its zeros inside the unit circle on the z-plane. To find this polynomial, we determine the $2M$ roots of the right-hand-side of $|A(f)|^2$ and keep only those M that lie inside the unit circle. We start with the equation for the roots:

$$\sigma^2\big[1 + e_1^2 T_M^2(x_0\cos\delta)\big] = 0 \quad \Rightarrow \quad T_M(x_0\cos\delta) = \pm\frac{j}{e_1}$$

Because $T_M(x_0\cos\delta) = \cos\big(M\,\text{acos}(x_0\cos\delta)\big)$, the desired M roots are given by:

$$x_0\cos\delta_m = \cos\left(\frac{\text{acos}\big(-\frac{j}{e_1}\big) + m\pi}{M}\right), \quad m = 0, 1, \dots, M-1 \tag{6.8.15}$$

Indeed, these satisfy:

$$\cos\big(M\,\text{acos}(x_0\cos\delta_m)\big) = \cos\left(\text{acos}\big(-\frac{j}{e_1}\big) + m\pi\right) = -\frac{j}{e_1}\cos m\pi = \pm\frac{j}{e_1}$$

Solving Eq. (6.8.15) for δ_m, we find:

$$\boxed{\delta_m = \text{acos}\left[\frac{1}{x_0}\cos\left(\frac{\text{acos}\big(-\frac{j}{e_1}\big) + m\pi}{M}\right)\right]}, \quad m = 0, 1, \dots, M-1 \tag{6.8.16}$$

Then, the M zeros of $A(z)$ are constructed by:

$$\boxed{z_m = e^{2j\delta_m}}, \quad m = 0, 1, \dots, M-1 \tag{6.8.17}$$

These zeros lie inside the unit circle, $|z_m| < 1$. (Replacing $-j/e_1$ by $+j/e_1$ in Eq. (6.8.16) would generate M zeros that lie outside the unit circle; these are the zeros of $\bar{A}(z)$.) Finally, the polynomial $A(z)$ is obtained by multiplying the root factors:

$$A(z) = \prod_{m=0}^{M-1}(1 - z_m z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_M z^{-M} \tag{6.8.18}$$

Once $A(z)$ is obtained, we may fix the scale factor σ^2 by requiring that the two sides of Eq. (6.8.14) match at $f = 0$. Noting that $A(f)$ at $f = 0$ is equal to the sum of the coefficients of $A(z)$ and that $e_1 T_M(x_0) = e_0$, we obtain the condition:

$$\left|\sum_{m=0}^{M-1} a_m\right|^2 = \sigma^2(1 + e_0^2) \quad \Rightarrow \quad \sigma = \pm\frac{\left|\sum_{m=0}^{M-1} a_m\right|}{\sqrt{1 + e_0^2}} \tag{6.8.19}$$

Either sign of σ leads to a solution, but its physical realizability (i.e., $n_1 \geq 1$) requires that we choose the negative sign if $n_a < n_b$, and the positive one if $n_a > n_b$. (The opposite choice of signs leads to the solution $n_i' = n_a^2/n_i$, $i = a, 1, \dots, M, b$.)

The polynomial $B(z)$ can now be constructed by taking the square root of the second equation in (6.8.14). Again, the simplest procedure is to determine the roots of the right-hand side and multiply the root factors. The root equations are:

$$\sigma^2 e_1^2 T_M^2(x_0\cos\delta) = 0 \quad \Rightarrow \quad T_M(x_0\cos\delta) = 0$$

with M roots:

$$\delta_m = \text{acos}\left(\frac{1}{x_0}\cos\big(\frac{(m+0.5)\pi}{M}\big)\right), \quad m = 0, 1, \dots, M-1 \tag{6.8.20}$$

The z-plane roots are $z_m = e^{2j\delta_m}$, $m = 0, 1, \dots, M-1$. The polynomial $B(z)$ is now constructed up to a constant b_0 by the product:

$$B(z) = b_0\prod_{m=0}^{M-1}(1 - z_m z^{-1}) \tag{6.8.21}$$

As before, the factor b_0 is fixed by matching Eq. (6.8.14) at $f = 0$. Because δ_m is real, the zeros z_m will all have unit magnitude and $B(z)$ will be equal to its reverse polynomial, $B^R(z) = B(z)$.

Finally, the reflection coefficients at the interfaces and the refractive indices are obtained by sending $A(z)$ and $B(z)$ into the backward layer recursion.

The above design steps are implemented by the MATLAB functions `chebtr`, `chebtr2`, and `chebtr3` with usage:

```
[n,a,b]    = chebtr(na,nb,A,DF);           % Chebyshev multilayer design
[n,a,b,A]  = chebtr2(na,nb,M,DF);          % specify order and bandwidth
[n,a,b,DF] = chebtr3(na,nb,M,A);           % specify order and attenuation
```

The inputs are the refractive indices n_a, n_b of the left and right media, the desired attenuation in dB, and the fractional bandwidth $\Delta F = \Delta f/f_0$. The output is the refractive index vector $\mathbf{n} = [n_a, n_1, n_2, \dots, n_M, n_b]$ and the reflection and transmission polynomials **b** and **a**. In `chebtr2` and `chebtr3`, the order M is given. To clarify the design steps, we give below the essential source code for `chebtr`:

```
e0 = sqrt((nb-na)^2/(4*nb*na));
x0 = 1/sin(DF*pi/4);
M = ceil(acosh(sqrt((e0^2+1)*10^(A/10) - e0^2))/acosh(x0));
```

```
e1 = e0/cosh(M*acosh(x0));

m=0:M-1;
delta = acos(cos((acos(-j/e1)+pi*m)/M)/x0);
z = exp(2*j*delta);                              % zeros of A(z)

a = real(poly(z));                               % coefficients of A(z)

sigma = sign(na-nb)*abs(sum(a))/sqrt(1+e0^2);    % scale factor σ

delta = acos(cos((m+0.5)*pi/M)/x0);
z = exp(2*j*delta);                              % zeros of B(z)

b = real(poly(z));                               % unscaled coefficients of B(z)
b0 = sigma * e0 / abs(sum(b));
b = b0 * b;                                      % rescaled B(z)

r = bkwrec(a,b);                                 % backward recursion
n = na * r2n(r);                                 % refractive indices
```

Example 6.8.1: *Broadband antireflection coating.* Design a broadband antireflection coating on glass with $n_a = 1$, $n_b = 1.5$, $A = 20$ dB, and fractional bandwidth $\Delta F = \Delta f/f_0 = 1.5$. Then, design a coating with deeper and narrower bandwidth having parameters $A = 30$ dB and $\Delta F = \Delta f/f_0 = 1.0$.

Solution: The reflectances of the designed coatings are shown in Fig. 6.8.2. The two cases have $M = 8$ and $M = 5$, respectively, and refractive indices:

$$n = [1, 1.0309, 1.0682, 1.1213, 1.1879, 1.2627, 1.3378, 1.4042, 1.4550, 1.5]$$

$$n = [1, 1.0284, 1.1029, 1.2247, 1.3600, 1.4585, 1.5]$$

The specifications are better than satisfied because the method rounds up the exact value of M to the next integer. These exact values were $M_{\text{exact}} = 7.474$ and $M_{\text{exact}} = 4.728$, and were increased to $M = 8$ and $M = 5$.

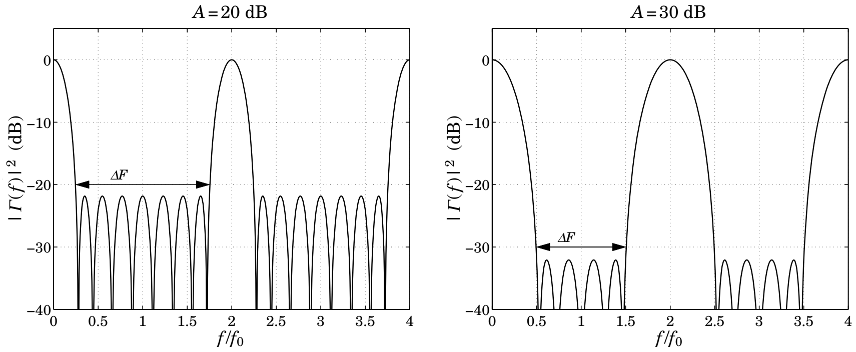


Fig. 6.8.2 Chebyshev designs. Reflectances are normalized to 0 dB at dc.

The desired bandedges shown on the graphs were computed from $f_1/f_0 = 1 - \Delta F/2$ and $f_1/f_0 = 1 + \Delta F/2$. The designed polynomial coefficients $\mathbf{a}, \mathbf{b}$ were in the two cases:

$$\mathbf{a} = \begin{bmatrix} 1.0000 \\ 0.0046 \\ 0.0041 \\ 0.0034 \\ 0.0025 \\ 0.0017 \\ 0.0011 \\ 0.0005 \\ 0.0002 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.0152 \\ -0.0178 \\ -0.0244 \\ -0.0290 \\ -0.0307 \\ -0.0290 \\ -0.0244 \\ -0.0178 \\ -0.0152 \end{bmatrix} \quad \text{and} \quad \mathbf{a} = \begin{bmatrix} 1.0000 \\ 0.0074 \\ 0.0051 \\ 0.0027 \\ 0.0010 \\ 0.0002 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.0140 \\ -0.0350 \\ -0.0526 \\ -0.0526 \\ -0.0350 \\ -0.0140 \end{bmatrix}$$

The zeros of the polynomials $\mathbf{a}$ were in the two cases:

$$z = \begin{bmatrix} 0.3978\angle \pm 27.93^o \\ 0.3517\angle \pm 73.75^o \\ 0.3266\angle \pm 158.76^o \\ 0.3331\angle \pm 116.34^o \end{bmatrix} \quad \text{and} \quad z = \begin{bmatrix} 0.2112\angle \pm 45.15^o \\ 0.1564\angle 180^o \\ 0.1678\angle \pm 116.30^o \end{bmatrix}$$

They lie inside the unit circle by design. The typical MATLAB code used to generate these examples was:

```
na = 1; nb = 1.5; A = 20; DF = 1.5;

n = chebtr(na,nb,A,DF);
M = length(n) - 2;

f = linspace(0,4,1601);
L = 0.25 * ones(1,M);

G0 = (na-nb)^2 / (na+nb)^2;
G  = abs(multidiel(n,L,1./f)).^2;

plot(f, 10*log10(G/G0));
```

The reflectances were computed with the function `multidiel`. The optical thickness inputs to `multidiel` were all quarter-wavelength at f_0. □

We note, in this example, that the coefficients of the polynomial $B(z)$ are symmetric about their middle, that is, the polynomial is self-reversing $B^R(z) = B(z)$. One consequence of this property is that the vector of reflection coefficients is also symmetric about its middle, that is,

$$[\rho_1, \rho_2, \dots, \rho_M, \rho_{M+1}] = [\rho_{M+1}, \rho_M, \dots, \rho_2, \rho_1] \tag{6.8.22}$$

or, $\rho_i = \rho_{M+2-i}$, for $i = 1, 2, \dots, M+1$. These conditions are equivalent to the following constraints among the resulting refractive indices:

$$\boxed{n_i n_{M+2-i} = n_a n_b \quad \Leftrightarrow \quad \rho_i = \rho_{M+2-i}}, \quad i = 1, 2, \dots, M+1 \tag{6.8.23}$$

These can be verified easily in the above example. The proof of these conditions follows from the symmetry of $B(z)$. A simple argument is to use the single-reflection approximation discussed in Example 6.6.4, in which the polynomial $B(z)$ is to first-order in the ρ_is:

$$B(z)=\rho_1+\rho_2 z^{-1}+\cdots+\rho_{M+1}z^{-M}$$

If the symmetry property $\rho_i=\rho_{M+2-i}$ were not true, then $B(z)$ could not satisfy the property $B^R(z)=B(z)$. A more exact argument that does not rely on this approximation can be given by considering the product of matrices (6.6.17).

In the design steps outlined above, we used MATLAB's built-in function `poly.m` to construct the numerator and denominator polynomials $B(z), A(z)$ from their zeros. These zeros are almost equally-spaced around the unit circle and get closer to each other with increasing order M. This causes `poly` to lose accuracy around order 50–60.

In the three `chebtr` functions (as well as in the Dolph-Chebyshev array functions of Chap. 20), we have used an improved version, `poly2.m`, with the same usage as `poly`, that maintains its accuracy up to order of about 3000.

Fig. 6.8.3 shows a typical pattern of zeros for Example 6.8.1 for normalized bandwidths of $\Delta F=1.85$ and $\Delta F=1.95$ and attenuation of $A=30$ dB. The zeros of $B(z)$ lie on the unit circle, and those of $A(z)$, inside the circle. The function `poly2` groups the zeros in subgroups such that the zeros within each subgroup are not as closely spaced. For example, for the left graph of Fig. 6.8.3, `poly2` picks the zeros sequentially, whereas for the right graph, it picks every other zero, thus forming two subgroups, then `poly` is called on each subgroup, and the two resulting polynomials are convolved to get the overall polynomial.

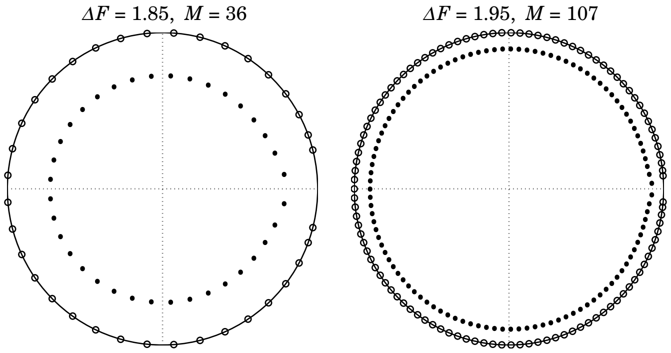


Fig. 6.8.3 Zero patterns of $B(z)$ (open circles) and $A(z)$ (filled circles), for $A=30$ dB.

Finally, we discuss the design of broadband terminations of transmission lines shown in Fig. 6.7.1. Because the media admittances are proportional to the refractive indices, $\eta_i^{-1}=n_i\eta_{\text{vac}}^{-1}$, we need only replace n_i by the line characteristic admittances:

$$[n_a,n_1,\dots,n_M,n_b]\to[Y_a,Y_1,\dots,Y_M,Y_b]$$

where Y_a, Y_b are the admittances of the main line and the load and Y_i, the admittances of the segments. Thus, the vector of admittances can be obtained by the MATLAB call:

```
Y = chebtr(Ya, Yb, A, DF);          % Chebyshev transmission line impedance transformer
```

We also have the property (6.8.23), $Y_iY_{M+2-i}=Y_aY_b$, or, $Z_iZ_{M+2-i}=Z_aZ_b$, for $i=1,2,\dots,M+1$, where $Y_i=1/Z_i$. One can work directly with impedances—the following call would generate exactly the same solution, where $Z=[Z_a,Z_1,\dots,Z_M,Z_b]$:

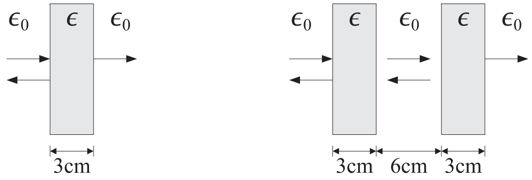
```
Z = chebtr(Za, Zb, A, DF);          % Chebyshev transmission line impedance transformer
```

In this design method, one does not have any control over the resulting refractive indices n_i or admittances Y_i. This can be problematic in the design of antireflection coatings because there do not necessarily exist materials with the designed n_is. However, one can replace or "simulate" any value of the refractive index of a layer by replacing the layer with an equivalent set of three layers of available indices and appropriate thicknesses [592–652].

This is not an issue in the case of transmission lines, especially microstrip lines, because one can design a line segment of a desired impedance by adjusting the geometry of the line, for example, by changing the diameters of a coaxial cable, the spacing of a parallel-wire, or the width of a microstrip line.

6.9 Problems

6.1 A uniform plane wave of frequency of 1.25 GHz is normally incident from free space onto a fiberglass dielectric slab ($\epsilon=4\epsilon_0, \mu=\mu_0$) of thickness of 3 cm, as shown on the left figure below.

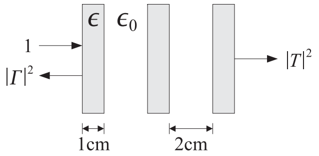


a. What is the free-space wavelength of this wave in cm? What is its wavelength inside the fiberglass?

b. What percentage of the incident power is reflected backwards?

c. Next, an identical slab is inserted to the right of the first slab at a distance of 6 cm, as shown on the right. What percentage of incident power is now reflected back?

6.2 Three identical dielectric slabs of thickness of 1 cm and dielectric constant $\epsilon=4\epsilon_0$ are positioned as shown below. A uniform plane wave of frequency of 3.75 GHz is incident normally onto the leftmost slab.



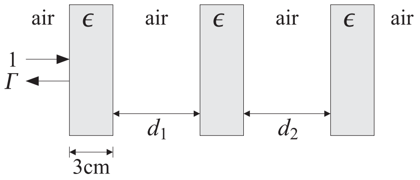
a. Determine the power reflection and transmission coefficients, $|\Gamma|^2$ and $|T|^2$, as percentages of the incident power.

b. Determine $|\Gamma|^2$ and $|T|^2$ if the three slabs and air gaps are replaced by a single slab of thickness of 7 cm.

6.3 Three identical fiberglass slabs of thickness of 3 cm and dielectric constant $\epsilon = 4\epsilon_0$ are positioned at separations $d_1 = d_2 = 6$ cm, as shown below. A wave of free-space wavelength of 24 cm is incident normally onto the left slab.

a. Determine the percentage of reflected power.

b. Repeat if the slabs are repositioned such that $d_1 = 12$ cm and $d_2 = 6$ cm.

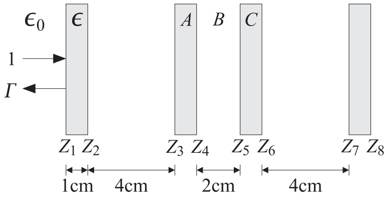


6.4 Four identical dielectric slabs of thickness of 1 cm and dielectric constant $\epsilon = 4\epsilon_0$ are positioned as shown below. A uniform plane wave of frequency of 3.75 GHz is incident normally onto the leftmost slab.

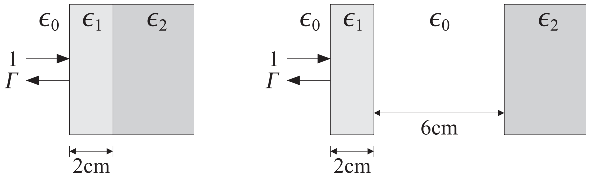
a. Determine the reflectance $|\Gamma|^2$ as a percentage.

b. Determine $|\Gamma|^2$ if slabs A and C are removed and replaced by air.

c. Determine $|\Gamma|^2$ if the air gap B between slabs A and C is filled with the same dielectric, so that ABC is a single slab.



6.5 A 2.5 GHz wave is normally incident from air onto a dielectric slab of thickness of 2 cm and refractive index of 1.5, as shown below. The medium to the right of the slab has index 2.25.



a. Derive an analytical expression of the reflectance $|\Gamma(f)|^2$ as a function of frequency and sketch it versus f over the interval $0 \le f \le 10$ GHz. What is the value of the reflectance at 2.5 GHz?

b. Next, the 2-cm slab is moved to the left by a distance of 6 cm, creating an air-gap between it and the rightmost dielectric. What is the value of the reflectance at 2.5 GHz?

6.6 Show that the antireflection coating design equations (6.2.2) can be written in the alternative forms:

$$\cos^2 k_2 l_2 = \frac{(n_2^2 - n_a n_b)\,(n_2^2 n_a - n_1^2 n_b)}{n_a\,(n_2^2 - n_b^2)\,(n_2^2 - n_1^2)}\,, \quad \sin^2 k_2 l_2 = \frac{n_2^2\,(n_b - n_a)\,(n_1^2 - n_a n_b)}{n_a\,(n_2^2 - n_b^2)\,(n_2^2 - n_1^2)}$$

Making the assumptions that $n_2 > n_1 > n_a$, $n_2 > n_b$, and $n_b > n_a$, show that for the design to have a solution, the following conditions must be satisfied:

$$n_1 > \sqrt{n_a n_b} \quad \text{and} \quad n_2 > n_1\sqrt{\frac{n_b}{n_a}}$$

6.7 Show that the characteristic polynomial of any 2×2 matrix F is expressible in terms of the trace and the determinant of F as in Eq. (6.3.10), that is,

$$\det(F - \lambda I) = \lambda^2 - (\operatorname{tr} F)\lambda + \det F$$

Moreover, for a unimodular matrix show that the two eigenvalues are $\lambda_\pm = e^{\pm\alpha}$ where $\alpha = \operatorname{acosh}(a)$ and $a = \operatorname{tr} F/2$.

6.8 Show that the bandedge condition $a = -1$ for a dielectric mirror is equivalent to the condition of Eq. (6.3.16). Moreover, show that an alternative condition is:

$$\cos\delta_H \cos\delta_L - \frac{1}{2}\left(\frac{n_H}{n_L} + \frac{n_L}{n_H}\right)\sin\delta_H \sin\delta_L = -1$$

6.9 Stating with the approximate bandedge frequencies given in Eq. (6.3.19), show that the bandwidth and center frequency of a dielectric mirror are given by:

$$\Delta f = f_2 - f_1 = \frac{2f_0\,\operatorname{asin}(\rho)}{\pi(L_H + L_L)}\,, \quad f_c = \frac{f_1 + f_2}{2} = \frac{f_0}{2(L_H + L_L)}$$

where $L_H = n_H l_H/\lambda_0$, $L_L = n_L l_L/\lambda_0$, and λ_0 is a normalization wavelength, and f_0 the corresponding frequency $f_0 = c_0/\lambda_0$.

6.10 *Computer Experiment—Antireflection Coatings.* Compute and plot over the 400–700 nm visible band the reflectance of the following antireflection coatings on glass, defined by the refractive indices and normalized optical thicknesses:

a. $n = [1, 1.38, 1.5]$, $L = [0.25]$
b. $n = [1, 1.38, 1.63, 1.5]$, $L = [0.25, 0.50]$
c. $n = [1, 1.38, 2.2, 1.63, 1.5]$, $L = [0.25, 0.50, 0.25]$
d. $n = [1, 1.38, 2.08, 1.38, 2.08, 1.5]$, $L = [0.25, 0.527, 0.0828, 0.0563]$

The normalization wavelength is $\lambda_0 = 550$ nm. Evaluate and compare the coatings in terms of bandwidth. Cases (a-c) are discussed in Sec. 6.2 and case (d) is from [599].

6.11 *Computer Experiment—Dielectric Sunglasses.* A thin-film multilayer design of dielectric sunglasses was carried out in Ref. [1321] using 29 layers of alternating TiO_2 ($n_H = 2.35$) and SiO_2 ($n_L = 1.45$) coating materials. The design may be found on the web page:
`www.sspectra.com/designs/sunglasses.html`.

The design specifications for the thin-film structure were that the transmittance be: (a) less than one percent for wavelengths 400–500 nm, (b) between 15–25 percent for 510–790 nm, and (c) less than one percent for 800–900 nm.

Starting with the high-index layer closest to the air side and ending with the high-index layer closest to the glass substrate, the designed lengths of the 29 layers were in nm (read across):

21.12	32.41	73.89	123.90	110.55	129.47
63.17	189.07	68.53	113.66	62.56	59.58
27.17	90.29	44.78	73.58	50.14	94.82
60.40	172.27	57.75	69.00	28.13	93.12
106.07	111.15	32.68	32.82	69.95	

Form the optical lengths $n_i l_i$ and normalize them $L_i = n_i l_i/\lambda_0$, such that the maximum optical length is a quarter wavelength at λ_0. What is the value of λ_0 in nm? Assuming the glass substrate has index $n = 1.5$, compute and plot the reflectance and transmittance over the band 400–900 nm.

6.12 *Computer Experiment—Dielectric Mirror.* Reproduce all the results and graphs of Example 6.3.2. In addition, carry out the computations for the cases of $N = 16, 32$ bilayers.

In all cases, calculate the minimum and maximum reflectance within the high-reflectance band. For one value of N, calculate the reflectance using the closed-form expression (6.3.33) and verify that it is the same as that produced by `multidiel`.

6.13 *Computer Experiment—Dielectric Mirror.* Reproduce all the results and graphs of Example 6.3.3. Repeat the computations and plots when the number of bilayers is $N = 8, 16$. Repeat for $N = 4, 8, 16$ assuming the layers are quarter-wavelength layers at 12.5 μm. In all cases, calculate the minimum and maximum reflectance within the high-reflectance band.

6.14 *Computer Experiment—Shortpass and Longpass Filters.* Reproduce all the results and graphs of Example 6.3.5. Redo the experiments by shifting the short-pass wavelength to $\lambda_0 = 750$ nm in the first case, and the long-pass wavelength to $\lambda_0 = 350$ nm in the second case. Plot the reflectances over the extended band of 200–1000 nm.

6.15 *Computer Experiment—Wide Infrared Bandpass Filter.* A 47-layer infrared bandpass filter with wide transmittance bandwidth was designed in Ref. [1321]. The design may be found on the web page `www.sspectra.com/designs/irbp.html`.

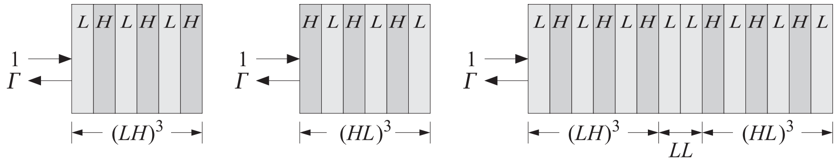
The alternating low- and high-index layers were ZnS and Ge with indices 2.2 and 4.2. The substrate was Ge with index 4. The design specifications were that the transmittance be: (a) less than 0.1% for wavelengths 2–3 μm, (b) greater than 99% for 3.3–5 μm, and (c) less than 0.1% for 5.5–7 μm.

Starting with a low-index layer near the air side and ending with a low-index layer at the substrate, the layer lengths were in nm (read across):

528.64	178.96	250.12	123.17	294.15	156.86	265.60	134.34
266.04	147.63	289.60	133.04	256.22	165.16	307.19	125.25
254.28	150.14	168.55	68.54	232.65	125.48	238.01	138.25
268.21	98.28	133.58	125.31	224.72	40.79	564.95	398.52
710.47	360.01	724.86	353.08	718.52	358.23	709.26	370.42
705.03	382.28	720.06	412.85	761.47	48.60	97.33	

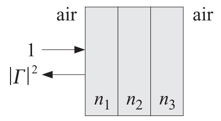
Form the optical lengths $n_i l_i$ and normalize them $L_i = n_i l_i/\lambda_0$, such that the maximum optical length is a quarter wavelength at λ_0. What is the value of λ_0 in μm? Compute and plot the reflectance and transmittance over the band 2–7 μm.

6.16 The figure below shows three multilayer structures. The first, denoted by $(LH)^3$, consists of three identical bilayers, each bilayer consisting of a low-index and a high-index quarter-wave layer, with indices $n_L = 1.38$ and $n_H = 3.45$. The second multilayer, denoted by $(HL)^3$, is the same as the first one, but with the order of the layers reversed. The third one, denoted by $(LH)^3(LL)(HL)^3$ consists of the first two side-by-side and separated by two low-index quarter-wave layers LL.



In all three cases, determine the overall reflection response Γ, as well as the percentage of reflected power, at the design frequency at which the individual layers are quarter-wave.

6.17 A radome protecting a microwave transmitter consists of a three-slab structure as shown below. The medium to the left and right of the structure is air. At the carrier frequency of the transmitter, the structure is required to be reflectionless, that is, $\Gamma = 0$.

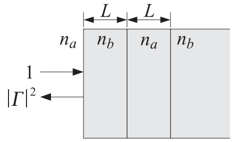


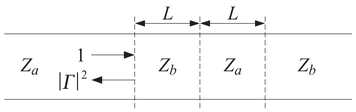
a. Assuming that all three slabs are quarter-wavelength at the design frequency, what should be the relationship among the three refractive indices n_1, n_2, n_3 in order to achieve a reflectionless structure?

b. What should be the relationship among the refractive indices n_1, n_2, n_3 if the middle slab (i.e., n_2) is half-wavelength but the other two are still quarter-wavelength slabs?

c. For case (a), suppose that the medium to the right has a slightly different refractive index from that of air, say, $n_b = 1+\epsilon$. Calculate the small resulting reflection response Γ to first order in ϵ.

6.18 In order to obtain a reflectionless interface between media n_a and n_b, two dielectric slabs of equal optical lengths L and refractive indices n_b, n_a are positioned as shown below. (The same technique can be used to connect two transmission lines of impedances Z_a and Z_b.)





A plane wave of frequency f is incident normally from medium n_a. Let f_0 be the frequency at which the structure must be reflectionless. Let L be the common optical length normalized to the free-space wavelength $\lambda_0 = c_0/f_0$, that is, $L = n_a l_a/\lambda_0 = n_b l_b/\lambda_0$.

a. Show that the reflection response into medium n_a is given by:

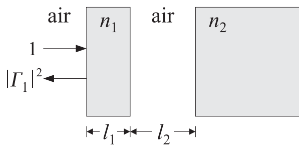
$$\Gamma = \rho\,\frac{1-(1+\rho^2)e^{-2j\delta}+e^{-4j\delta}}{1-2\rho^2 e^{-2j\delta}+\rho^2 e^{-4j\delta}}\,,\qquad \rho = \frac{n_a-n_b}{n_a+n_b}\,,\qquad \delta = 2\pi L\frac{f}{f_0}$$

b. Show that the interface will be reflectionless at frequency f_0 provided the optical lengths are chosen according to:

$$L = \frac{1}{4\pi}\arccos\left(\frac{1+\rho^2}{2}\right)$$

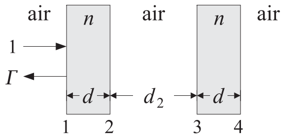
This is known as a *twelfth-wave* transformer because for $\rho = 0$, it gives $L = 1/12$.

6.19 A lossless dielectric slab of refractive index n_1 and thickness l_1 is positioned at a distance l_2 from a semi-infinite dielectric of refractive index n_2, as shown below.



A uniform plane wave of free-space wavelength λ_0 is incident normally on the slab from the left. Assuming that the slab n_1 is a quarter-wavelength slab, determine the length l_2 (in units of λ_0) and the relationship between n_1 and n_2 in order that there be no reflected wave into the leftmost medium (i.e., $\Gamma_1 = 0$).

6.20 In order to provide structural strength and thermal insulation, a radome is constructed using two identical dielectric slabs of length d and refractive index n, separated by an air-gap of length d_2, as shown below.



Recall that a reflectionless single-layer radome requires that the dielectric layer have half-wavelength thickness.

However, show that for the above dual-slab arrangement, either half- or quarter-wavelength dielectric slabs may be used, provided that the middle air-gap is chosen to be a half-wavelength layer, i.e., $d_2 = \lambda_0/2$, at the operating wavelength λ_0. [*Hint:* Work with wave impedances at the operating wavelength.]

6.21 *Computer Experiment—Dielectric Mirror Bands.* Consider the trace function given by Eq. (6.3.13) of the text, that is,

$$a = \frac{\cos(\delta_H + \delta_L) - \rho^2\cos(\delta_H - \delta_L)}{1 - \rho^2}$$

The purpose of this problem is to study a as a function of frequency, which enters through:

$$\delta_i = 2\pi\left(\frac{f}{f_0}\right)L_i\,, \quad L_i = \frac{n_i l_i}{\lambda_0}\,, \quad i = H, L$$

and to identify the frequency bands where a switches from $|a| \le 1$ to $|a| \ge 1$, that is, when the dielectric mirror structure switches from transmitting to reflecting.

a. For the parameters given in Example 6.3.2 of the text, make a plot of a versus f over the range $0 \le f \le 4f_0$, using f/f_0 as your x-axis. Place on the graph the left and right bandedge frequencies f_1, f_2 of the reflecting bands centered at f_0 and odd multiples thereof.

b. Repeat for the parameters $n_a = 1$, $n_H = 4.6$, $n_L = 1.6$, $L_H = 0.3$, $L_L = 0.2$. These parameters are close to those of Example 6.3.2. You may use the function `omniband` to calculate the left and right bandedge frequencies around f_0.

In plotting a versus f/f_0, you will notice that a can become greater than +1 near $f = 2f_0$. Determine the left and right bandedge frequencies around $2f_0$ and check to see whether they define another reflecting band around $2f_0$.

7

Oblique Incidence

7.1 Oblique Incidence and Snel's Laws

With some redefinitions, the formalism of transfer matrices and wave impedances for normal incidence translates almost verbatim to the case of oblique incidence.

By separating the fields into *transverse* and longitudinal components with respect to the direction the dielectrics are stacked (the z-direction), we show that the transverse components satisfy the identical transfer matrix relationships as in the case of normal incidence, provided we replace the media impedances η by the *transverse impedances* η_T defined below.

Fig. 7.1.1 depicts plane waves incident from both sides onto a planar interface separating two media ϵ, ϵ'. Both cases of parallel and perpendicular polarizations are shown.

In *parallel* polarization, also known as p-polarization, π-polarization, or TM polarization, the electric fields lie *on the plane* of incidence and the magnetic fields are

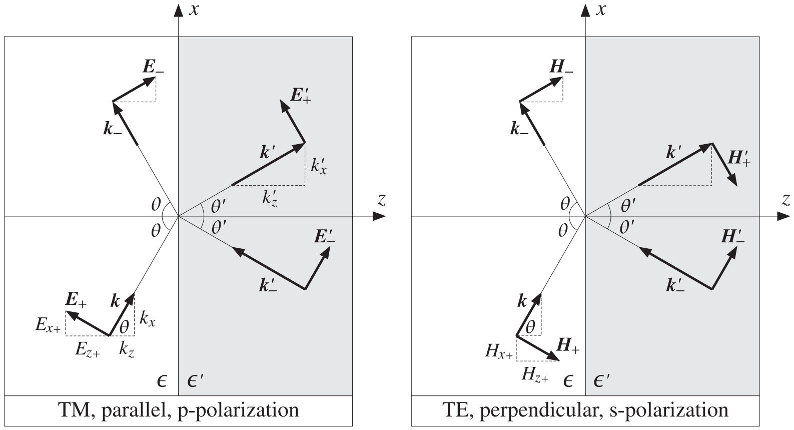


Fig. 7.1.1 Oblique incidence for TM- and TE-polarized waves.

perpendicular to that plane (along the y-direction) and *transverse* to the z-direction.

In *perpendicular* polarization, also known as s-polarization,[†] σ-polarization, or TE polarization, the electric fields are perpendicular to the plane of incidence (along the y-direction) and *transverse* to the z-direction, and the magnetic fields lie on that plane.

The figure shows the angles of incidence and reflection to be the same on either side. This is Snel's law[†] of reflection and is a consequence of the boundary conditions.

The figure also implies that the two planes of incidence and two planes of reflection all coincide with the xz-plane. This is also a consequence of the boundary conditions.

Starting with arbitrary wavevectors $\mathbf{k}_\pm = \hat{\mathbf{x}}\,k_{x\pm} + \hat{\mathbf{y}}\,k_{y\pm} + \hat{\mathbf{z}}\,k_{z\pm}$ and similarly for $\mathbf{k}'_\pm$, the incident and reflected electric fields at the two sides will have the general forms:

$$\mathbf{E}_+ e^{-j\mathbf{k}_+\cdot\mathbf{r}}\,, \quad \mathbf{E}_- e^{-j\mathbf{k}_-\cdot\mathbf{r}}\,, \quad \mathbf{E}'_+ e^{-j\mathbf{k}'_+\cdot\mathbf{r}}\,, \quad \mathbf{E}'_- e^{-j\mathbf{k}'_-\cdot\mathbf{r}}$$

The boundary conditions state that the net *transverse* (tangential) component of the electric field must be continuous across the interface. Assuming that the interface is at $z = 0$, we can write this condition in a form that applies to both polarizations:

$$\mathbf{E}_{T+}e^{-j\mathbf{k}_+\cdot\mathbf{r}} + \mathbf{E}_{T-}e^{-j\mathbf{k}_-\cdot\mathbf{r}} = \mathbf{E}'_{T+}e^{-j\mathbf{k}'_+\cdot\mathbf{r}} + \mathbf{E}'_{T-}e^{-j\mathbf{k}'_-\cdot\mathbf{r}}\,, \qquad \text{at} \quad z = 0 \tag{7.1.1}$$

where the subscript T denotes the transverse (with respect to z) part of a vector, that is, $\mathbf{E}_T = \hat{\mathbf{z}}\times(\mathbf{E}\times\hat{\mathbf{z}}) = \mathbf{E} - \hat{\mathbf{z}}\,E_z$. Setting $z = 0$ in the propagation phase factors, we obtain:

$$\mathbf{E}_{T+}e^{-j(k_{x+}x+k_{y+}y)} + \mathbf{E}_{T-}e^{-j(k_{x-}x+k_{y-}y)} = \mathbf{E}'_{T+}e^{-j(k'_{x+}x+k'_{y+}y)} + \mathbf{E}'_{T-}e^{-j(k'_{x-}x+k'_{y-}y)} \tag{7.1.2}$$

For the two sides to match at *all* points on the interface, the phase factors must be equal to each other for all x and y:

$$e^{-j(k_{x+}x+k_{y+}y)} = e^{-j(k_{x-}x+k_{y-}y)} = e^{-j(k'_{x+}x+k'_{y+}y)} = e^{-j(k'_{x-}x+k'_{y-}y)} \qquad \text{(phase matching)}$$

and this requires the x- and y-components of the wave vectors to be equal:

$$\begin{aligned} k_{x+} &= k_{x-} = k'_{x+} = k'_{x-} \\ k_{y+} &= k_{y-} = k'_{y+} = k'_{y-} \end{aligned} \tag{7.1.3}$$

If the left plane of incidence is the xz-plane, so that $k_{y+} = 0$, then all y-components of the wavevectors will be zero, implying that all planes of incidence and reflection will coincide with the xz-plane. In terms of the incident and reflected angles $\theta_\pm, \theta'_\pm$, the conditions on the x-components read:

$$k\sin\theta_+ = k\sin\theta_- = k'\sin\theta'_+ = k'\sin\theta'_- \tag{7.1.4}$$

These imply *Snel's law of reflection*:

$$\boxed{\begin{aligned} \theta_+ &= \theta_- \equiv \theta \\ \theta'_+ &= \theta'_- \equiv \theta' \end{aligned}} \qquad \text{(Snel's law of reflection)} \tag{7.1.5}$$

[†] from the German word *senkrecht* for perpendicular.

[†] named after Willebrord Snel, b.1580, almost universally misspelled as Snell.

And also *Snel's law of refraction*, that is, $k\sin\theta = k'\sin\theta'$. Setting $k = nk_0$, $k' = n'k_0$, and $k_0 = \omega/c_0$, we have:

$$\boxed{n\sin\theta = n'\sin\theta'} \quad \Rightarrow \quad \boxed{\frac{\sin\theta}{\sin\theta'} = \frac{n'}{n}} \qquad \text{(Snel's law of refraction)} \tag{7.1.6}$$

It follows that the wave vectors shown in Fig. 7.1.1 will be explicitly:

$$\begin{aligned} \mathbf{k} = \mathbf{k}_+ &= k_x\hat{\mathbf{x}} + k_z\hat{\mathbf{z}} = k\sin\theta\,\hat{\mathbf{x}} + k\cos\theta\,\hat{\mathbf{z}} \\ \mathbf{k}_- &= k_x\hat{\mathbf{x}} - k_z\hat{\mathbf{z}} = k\sin\theta\,\hat{\mathbf{x}} - k\cos\theta\,\hat{\mathbf{z}} \\ \mathbf{k}' = \mathbf{k}'_+ &= k'_x\hat{\mathbf{x}} + k'_z\hat{\mathbf{z}} = k'\sin\theta'\,\hat{\mathbf{x}} + k'\cos\theta'\,\hat{\mathbf{z}} \\ \mathbf{k}'_- &= k'_x\hat{\mathbf{x}} - k'_z\hat{\mathbf{z}} = k'\sin\theta'\,\hat{\mathbf{x}} - k'\cos\theta'\,\hat{\mathbf{z}} \end{aligned} \tag{7.1.7}$$

The net transverse electric fields at arbitrary locations on either side of the interface are given by Eq. (7.1.1). Using Eq. (7.1.7), we have:

$$\begin{aligned} \mathbf{E}_T(x,z) &= \mathbf{E}_{T+}e^{-j\mathbf{k}_+\cdot\mathbf{r}} + \mathbf{E}_{T-}e^{-j\mathbf{k}_-\cdot\mathbf{r}} = \big(\mathbf{E}_{T+}e^{-jk_zz} + \mathbf{E}_{T-}e^{jk_zz}\big)e^{-jk_xx} \\ \mathbf{E}'_T(x,z) &= \mathbf{E}'_{T+}e^{-j\mathbf{k}'_+\cdot\mathbf{r}} + \mathbf{E}'_{T-}e^{-j\mathbf{k}'_-\cdot\mathbf{r}} = \big(\mathbf{E}'_{T+}e^{-jk'_zz} + \mathbf{E}'_{T-}e^{jk'_zz}\big)e^{-jk'_xx} \end{aligned} \tag{7.1.8}$$

In analyzing multilayer dielectrics stacked along the z-direction, the phase factor $e^{-jk_xx} = e^{-jk'_xx}$ will be common at all interfaces, and therefore, we can ignore it and restore it at the end of the calculations, if so desired. Thus, we write Eq. (7.1.8) as:

$$\begin{aligned} \mathbf{E}_T(z) &= \mathbf{E}_{T+}e^{-jk_zz} + \mathbf{E}_{T-}e^{jk_zz} \\ \mathbf{E}'_T(z) &= \mathbf{E}'_{T+}e^{-jk'_zz} + \mathbf{E}'_{T-}e^{jk'_zz} \end{aligned} \tag{7.1.9}$$

In the next section, we work out explicit expressions for Eq. (7.1.9)

7.2 Transverse Impedance

The transverse components of the electric fields are defined differently in the two polarization cases. We recall from Sec. 2.9 that an obliquely-moving wave will have, in general, both TM and TE components. For example, according to Eq. (2.9.9), the wave incident on the interface from the left will be given by:

$$\begin{aligned} \mathbf{E}_+(\mathbf{r}) &= \big[(\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta)A_+ + \hat{\mathbf{y}}\,B_+\big]e^{-j\mathbf{k}_+\cdot\mathbf{r}} \\ \mathbf{H}_+(\mathbf{r}) &= \frac{1}{\eta}\big[\hat{\mathbf{y}}\,A_+ - (\hat{\mathbf{x}}\cos\theta - \hat{\mathbf{z}}\sin\theta)B_+\big]e^{-j\mathbf{k}_+\cdot\mathbf{r}} \end{aligned} \tag{7.2.1}$$

where the A_+ and B_+ terms represent the TM and TE components, respectively. Thus, the transverse components are:

$$\begin{aligned} \mathbf{E}_{T+}(x,z) &= \big[\hat{\mathbf{x}}\,A_+\cos\theta + \hat{\mathbf{y}}\,B_+\big]e^{-j(k_xx+k_zz)} \\ \mathbf{H}_{T+}(x,z) &= \frac{1}{\eta}\big[\hat{\mathbf{y}}\,A_+ - \hat{\mathbf{x}}\,B_+\cos\theta\big]e^{-j(k_xx+k_zz)} \end{aligned} \tag{7.2.2}$$

Similarly, the wave reflected back into the left medium will have the form:

$$\begin{aligned}
\boldsymbol{E}_-(\mathbf{r}) &= \big[(\hat{\mathbf{x}}\cos\theta+\hat{\mathbf{z}}\sin\theta)A_- + \hat{\mathbf{y}}B_-\big]e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}\\
\boldsymbol{H}_-(\mathbf{r}) &= \frac{1}{\eta}\big[-\hat{\mathbf{y}}A_- + (\hat{\mathbf{x}}\cos\theta+\hat{\mathbf{z}}\sin\theta)B_-\big]e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}
\end{aligned}\tag{7.2.3}$$

with corresponding transverse parts:

$$\begin{aligned}
\boldsymbol{E}_{T-}(x,z) &= \big[\hat{\mathbf{x}}A_-\cos\theta + \hat{\mathbf{y}}B_-\big]e^{-j(k_xx-k_zz)}\\
\boldsymbol{H}_{T-}(x,z) &= \frac{1}{\eta}\big[-\hat{\mathbf{y}}A_- + \hat{\mathbf{x}}B_-\cos\theta\big]e^{-j(k_xx-k_zz)}
\end{aligned}\tag{7.2.4}$$

Defining the transverse amplitudes and transverse impedances by:

$$\begin{aligned}
&A_{T\pm} = A_\pm\cos\theta\,, \qquad B_{T\pm} = B_\pm\\
&\eta_{TM} = \eta\cos\theta\,, \qquad \eta_{TE} = \frac{\eta}{\cos\theta}
\end{aligned}\tag{7.2.5}$$

and noting that $A_{T\pm}/\eta_{TM} = A_\pm/\eta$ and $B_{T\pm}/\eta_{TE} = B_\pm\cos\theta/\eta$, we may write Eq. (7.2.2) in terms of the transverse quantities as follows:

$$\begin{aligned}
\boldsymbol{E}_{T+}(x,z) &= \big[\hat{\mathbf{x}}A_{T+} + \hat{\mathbf{y}}B_{T+}\big]e^{-j(k_xx+k_zz)}\\
\boldsymbol{H}_{T+}(x,z) &= \Big[\hat{\mathbf{y}}\frac{A_{T+}}{\eta_{TM}} - \hat{\mathbf{x}}\frac{B_{T+}}{\eta_{TE}}\Big]e^{-j(k_xx+k_zz)}
\end{aligned}\tag{7.2.6}$$

Similarly, Eq. (7.2.4) is expressed as:

$$\begin{aligned}
\boldsymbol{E}_{T-}(x,z) &= \big[\hat{\mathbf{x}}A_{T-} + \hat{\mathbf{y}}B_{T-}\big]e^{-j(k_xx-k_zz)}\\
\boldsymbol{H}_{T-}(x,z) &= \Big[-\hat{\mathbf{y}}\frac{A_{T-}}{\eta_{TM}} + \hat{\mathbf{x}}\frac{B_{T-}}{\eta_{TE}}\Big]e^{-j(k_xx-k_zz)}
\end{aligned}\tag{7.2.7}$$

Adding up Eqs. (7.2.6) and (7.2.7) and ignoring the common factor e^{-jk_xx}, we find for the net transverse fields on the left side:

$$\begin{aligned}
\boldsymbol{E}_T(z) &= \hat{\mathbf{x}}E_{TM}(z) + \hat{\mathbf{y}}E_{TE}(z)\\
\boldsymbol{H}_T(z) &= \hat{\mathbf{y}}H_{TM}(z) - \hat{\mathbf{x}}H_{TE}(z)
\end{aligned}\tag{7.2.8}$$

where the TM and TE components have the same structure provided one uses the appropriate transverse impedance:

$$\begin{aligned}
E_{TM}(z) &= A_{T+}e^{-jk_zz} + A_{T-}e^{jk_zz}\\
H_{TM}(z) &= \frac{1}{\eta_{TM}}\big[A_{T+}e^{-jk_zz} - A_{T-}e^{jk_zz}\big]
\end{aligned}\tag{7.2.9}$$

$$\begin{aligned}
E_{TE}(z) &= B_{T+}e^{-jk_zz} + B_{T-}e^{jk_zz}\\
H_{TE}(z) &= \frac{1}{\eta_{TE}}\big[B_{T+}e^{-jk_zz} - B_{T-}e^{jk_zz}\big]
\end{aligned}\tag{7.2.10}$$

We summarize these in the compact form, where E_T stands for either E_{TM} or E_{TE}:

$$\boxed{\begin{aligned}
E_T(z) &= E_{T+}e^{-jk_zz} + E_{T-}e^{jk_zz}\\
H_T(z) &= \frac{1}{\eta_T}\big[E_{T+}e^{-jk_zz} - E_{T-}e^{jk_zz}\big]
\end{aligned}}\tag{7.2.11}$$

The transverse impedance η_T stands for either η_{TM} or η_{TE}:

$$\boxed{\eta_T = \begin{cases} \eta\cos\theta\,, & \text{TM, parallel, p-polarization}\\ \dfrac{\eta}{\cos\theta}\,, & \text{TE, perpendicular, s-polarization}\end{cases}}\tag{7.2.12}$$

Because $\eta = \eta_o/n$, it is convenient to define also a *transverse refractive index* through the relationship $\eta_T = \eta_0/n_T$. Thus, we have:

$$\boxed{n_T = \begin{cases} \dfrac{n}{\cos\theta}\,, & \text{TM, parallel, p-polarization}\\ n\cos\theta\,, & \text{TE, perpendicular, s-polarization}\end{cases}}\tag{7.2.13}$$

For the right side of the interface, we obtain similar expressions:

$$\begin{aligned}
E'_T(z) &= E'_{T+}e^{-jk'_zz} + E'_{T-}e^{jk'_zz}\\
H'_T(z) &= \frac{1}{\eta'_T}\big(E'_{T+}e^{-jk'_zz} - E'_{T-}e^{jk'_zz}\big)
\end{aligned}\tag{7.2.14}$$

$$\eta'_T = \begin{cases} \eta'\cos\theta'\,, & \text{TM, parallel, p-polarization}\\ \dfrac{\eta'}{\cos\theta'}\,, & \text{TE, perpendicular, s-polarization}\end{cases}\tag{7.2.15}$$

$$n'_T = \begin{cases} \dfrac{n'}{\cos\theta'}\,, & \text{TM, parallel, p-polarization}\\ n'\cos\theta'\,, & \text{TE, perpendicular, s-polarization}\end{cases}\tag{7.2.16}$$

where $E'_{T\pm}$ stands for $A'_{T\pm} = A'_\pm\cos\theta'$ or $B'_{T\pm} = B'_\pm$.

For completeness, we give below the complete expressions for the fields on both sides of the interface obtained by adding Eqs. (7.2.1) and (7.2.3), with all the propagation factors restored. On the left side, we have:

$$\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= \boldsymbol{E}_{TM}(\mathbf{r}) + \boldsymbol{E}_{TE}(\mathbf{r})\\
\boldsymbol{H}(\mathbf{r}) &= \boldsymbol{H}_{TM}(\mathbf{r}) + \boldsymbol{H}_{TE}(\mathbf{r})
\end{aligned}\tag{7.2.17}$$

where

$$\begin{aligned}
\boldsymbol{E}_{TM}(\mathbf{r}) &= (\hat{\mathbf{x}}\cos\theta-\hat{\mathbf{z}}\sin\theta)A_+e^{-j\,\mathbf{k}_+\cdot\mathbf{r}} + (\hat{\mathbf{x}}\cos\theta+\hat{\mathbf{z}}\sin\theta)A_-e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}\\
\boldsymbol{H}_{TM}(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{1}{\eta}\big(A_+e^{-j\,\mathbf{k}_+\cdot\mathbf{r}} - A_-e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}\big)\\
\boldsymbol{E}_{TE}(\mathbf{r}) &= \hat{\mathbf{y}}\,\big(B_+e^{-j\,\mathbf{k}_+\cdot\mathbf{r}} + B_-e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}\big)\\
\boldsymbol{H}_{TE}(\mathbf{r}) &= \frac{1}{\eta}\big[-(\hat{\mathbf{x}}\cos\theta-\hat{\mathbf{z}}\sin\theta)B_+e^{-j\,\mathbf{k}_+\cdot\mathbf{r}} + (\hat{\mathbf{x}}\cos\theta+\hat{\mathbf{z}}\sin\theta)B_-e^{-j\,\mathbf{k}_-\cdot\mathbf{r}}\big]
\end{aligned}\tag{7.2.18}$$

The transverse parts of these are the same as those given in Eqs. (7.2.9) and (7.2.10). On the right side of the interface, we have:

$$\begin{aligned} \boldsymbol{E}'(\mathbf{r}) &= \boldsymbol{E}'_{TM}(\mathbf{r})+\boldsymbol{E}'_{TE}(\mathbf{r}) \\ \boldsymbol{H}'(\mathbf{r}) &= \boldsymbol{H}'_{TM}(\mathbf{r})+\boldsymbol{H}'_{TE}(\mathbf{r}) \end{aligned} \tag{7.2.19}$$

$$\begin{aligned} \boldsymbol{E}'_{TM}(\mathbf{r}) &= (\hat{\mathbf{x}}\cos\theta' - \hat{\mathbf{z}}\sin\theta')\,A'_+ e^{-j\,\mathbf{k}'_+\cdot\mathbf{r}} + (\hat{\mathbf{x}}\cos\theta' + \hat{\mathbf{z}}\sin\theta')\,A'_- e^{-j\,\mathbf{k}'_-\cdot\mathbf{r}} \\ \boldsymbol{H}'_{TM}(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{1}{\eta'}\,(A'_+ e^{-j\,\mathbf{k}'_+\cdot\mathbf{r}} - A'_- e^{-j\,\mathbf{k}'_-\cdot\mathbf{r}}) \\ \boldsymbol{E}'_{TE}(\mathbf{r}) &= \hat{\mathbf{y}}\,(B'_+ e^{-j\,\mathbf{k}'_+\cdot\mathbf{r}} + B'_- e^{-j\,\mathbf{k}'_-\cdot\mathbf{r}}) \\ \boldsymbol{H}'_{TE}(\mathbf{r}) &= \frac{1}{\eta'}\big[-(\hat{\mathbf{x}}\cos\theta' - \hat{\mathbf{z}}\sin\theta')\,B'_+ e^{-j\,\mathbf{k}'_+\cdot\mathbf{r}} + (\hat{\mathbf{x}}\cos\theta' + \hat{\mathbf{z}}\sin\theta')\,B'_- e^{-j\,\mathbf{k}'_-\cdot\mathbf{r}}\big] \end{aligned} \tag{7.2.20}$$

7.3 Propagation and Matching of Transverse Fields

Eq. (7.2.11) has the identical form of Eq. (5.1.1) of the normal incidence case, but with the substitutions:

$$\boxed{\eta \to \eta_T\,, \quad e^{\pm jkz} \to e^{\pm jk_z z} = e^{\pm jkz\cos\theta}} \tag{7.3.1}$$

Every definition and concept of Chap. 5 translates into the oblique case. For example, we can define the *transverse wave impedance* at position z by:

$$Z_T(z) = \frac{E_T(z)}{H_T(z)} = \eta_T\,\frac{E_{T+}e^{-jk_z z} + E_{T-}e^{jk_z z}}{E_{T+}e^{-jk_z z} - E_{T-}e^{jk_z z}} \tag{7.3.2}$$

and the *transverse reflection coefficient* at position z:

$$\Gamma_T(z) = \frac{E_{T-}(z)}{E_{T+}(z)} = \frac{E_{T-}e^{jk_z z}}{E_{T+}e^{-jk_z z}} = \Gamma_T(0)\,e^{2jk_z z} \tag{7.3.3}$$

They are related as in Eq. (5.1.7):

$$Z_T(z) = \eta_T\,\frac{1+\Gamma_T(z)}{1-\Gamma_T(z)} \quad\Leftrightarrow\quad \Gamma_T(z) = \frac{Z_T(z)-\eta_T}{Z_T(z)+\eta_T} \tag{7.3.4}$$

The propagation matrices, Eqs. (5.1.11) and (5.1.13), relating the fields at two positions z_1, z_2 within the same medium, read now:

$$\begin{bmatrix} E_{T1+} \\ E_{T1-} \end{bmatrix} = \begin{bmatrix} e^{jk_z l} & 0 \\ 0 & e^{-jk_z l} \end{bmatrix} \begin{bmatrix} E_{T2+} \\ E_{T2-} \end{bmatrix} \qquad \text{(propagation matrix)} \tag{7.3.5}$$

$$\begin{bmatrix} E_{T1} \\ H_{T1} \end{bmatrix} = \begin{bmatrix} \cos k_z l & j\eta_T\sin k_z l \\ j\eta_T^{-1}\sin k_z l & \cos k_z l \end{bmatrix} \begin{bmatrix} E_{T2} \\ H_{T2} \end{bmatrix} \qquad \text{(propagation matrix)} \tag{7.3.6}$$

where $l = z_2 - z_1$. Similarly, the reflection coefficients and wave impedances propagate as:

$$\Gamma_{T1} = \Gamma_{T2}e^{-2jk_z l}\,, \qquad Z_{T1} = \eta_T\,\frac{Z_{T2} + j\eta_T\tan k_z l}{\eta_T + jZ_{T2}\tan k_z l} \tag{7.3.7}$$

The phase thickness $\delta = kl = 2\pi(nl)/\lambda$ of the normal incidence case, where λ is the free-space wavelength, is replaced now by:

$$\boxed{\delta_z = k_z l = kl\cos\theta = \frac{2\pi}{\lambda}\,nl\,\cos\theta} \tag{7.3.8}$$

At the interface $z = 0$, the boundary conditions for the tangential electric and magnetic fields give rise to the same conditions as Eqs. (5.2.1) and (5.2.2):

$$E_T = E'_T\,, \quad H_T = H'_T \tag{7.3.9}$$

and in terms of the forward/backward fields:

$$\begin{aligned} E_{T+} + E_{T-} &= E'_{T+} + E'_{T-} \\ \frac{1}{\eta_T}\,(E_{T+} - E_{T-}) &= \frac{1}{\eta'_T}\,(E'_{T+} - E'_{T-}) \end{aligned} \tag{7.3.10}$$

which can be solved to give the *matching matrix*:

$$\begin{bmatrix} E_{T+} \\ E_{T-} \end{bmatrix} = \frac{1}{\tau_T}\begin{bmatrix} 1 & \rho_T \\ \rho_T & 1 \end{bmatrix}\begin{bmatrix} E'_{T+} \\ E'_{T-} \end{bmatrix} \qquad \text{(matching matrix)} \tag{7.3.11}$$

where ρ_T, τ_T are transverse reflection coefficients, replacing Eq. (5.2.5):

$$\boxed{\begin{aligned} \rho_T &= \frac{\eta'_T - \eta_T}{\eta'_T + \eta_T} = \frac{n_T - n'_T}{n_T + n'_T} \\ \tau_T &= \frac{2\eta'_T}{\eta'_T + \eta_T} = \frac{2n_T}{n_T + n'_T} \end{aligned}} \qquad \text{(Fresnel coefficients)} \tag{7.3.12}$$

where $\tau_T = 1 + \rho_T$. We may also define the reflection coefficients from the right side of the interface: $\rho'_T = -\rho_T$ and $\tau'_T = 1 + \rho'_T = 1 - \rho_T$. Eqs. (7.3.12) are known as the *Fresnel* reflection and transmission coefficients.

The matching conditions for the transverse fields translate into corresponding matching conditions for the wave impedances and reflection responses:

$$Z_T = Z'_T \quad\Leftrightarrow\quad \Gamma_T = \frac{\rho_T + \Gamma'_T}{1 + \rho_T\Gamma'_T} \quad\Leftrightarrow\quad \Gamma'_T = \frac{\rho'_T + \Gamma_T}{1 + \rho'_T\Gamma_T} \tag{7.3.13}$$

If there is no left-incident wave from the right, that is, $E'_- = 0$, then, Eq. (7.3.11) takes the specialized form:

$$\begin{bmatrix} E_{T+} \\ E_{T-} \end{bmatrix} = \frac{1}{\tau_T}\begin{bmatrix} 1 & \rho_T \\ \rho_T & 1 \end{bmatrix}\begin{bmatrix} E'_{T+} \\ 0 \end{bmatrix} \tag{7.3.14}$$

which explains the meaning of the transverse reflection and transmission coefficients:

$$\rho_T = \frac{E_{T-}}{E_{T+}}\,, \quad \tau_T = \frac{E'_{T+}}{E_{T+}} \tag{7.3.15}$$

The relationship of these coefficients to the reflection and transmission coefficients of the *total* field amplitudes depends on the polarization. For TM, we have $E_{T\pm} = A_\pm \cos\theta$ and $E'_{T\pm} = A'_\pm \cos\theta'$, and for TE, $E_{T\pm} = B_\pm$ and $E'_{T\pm} = B'_\pm$. For both cases, it follows that the reflection coefficient ρ_T measures also the reflection of the total amplitudes, that is,

$$\rho_{TM} = \frac{A_- \cos\theta}{A_+ \cos\theta} = \frac{A_-}{A_+}\,, \qquad \rho_{TE} = \frac{B_-}{B_+}$$

whereas for the transmission coefficients, we have:

$$\tau_{TM} = \frac{A'_+ \cos\theta'}{A_+ \cos\theta} = \frac{\cos\theta'}{\cos\theta}\,\frac{A'_+}{A_+}\,, \qquad \tau_{TE} = \frac{B'_+}{B_+}$$

In addition to the boundary conditions of the transverse field components, there are also applicable boundary conditions for the longitudinal components. For example, in the TM case, the component E_z is normal to the surface and therefore, we must have the continuity condition $D_z = D'_z$, or $\epsilon E_z = \epsilon' E'_z$. Similarly, in the TE case, we must have $B_z = B'_z$. It can be verified that these conditions are automatically satisfied due to Snel's law (7.1.6).

The fields carry energy towards the z-direction, as well as the transverse x-direction. The energy flux along the z-direction must be conserved across the interface. The corresponding components of the Poynting vector are:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}\big[E_x H_y^* - E_y H_x^*\big]\,, \quad \mathcal{P}_x = \frac{1}{2}\,\mathrm{Re}\big[E_y H_z^* - E_z H_y^*\big]$$

For TM, we have $\mathcal{P}_z = \mathrm{Re}[E_x H_y^*]/2$ and for TE, $\mathcal{P}_z = -\mathrm{Re}[E_y H_x^*]/2$. Using the above equations for the fields, we find that $\mathcal{P}_z$ is given by the *same* expression for both TM and TE polarizations:

$$\mathcal{P}_z = \frac{\cos\theta}{2\eta}\big(|A_+|^2 - |A_-|^2\big)\,, \quad \text{or,} \quad \frac{\cos\theta}{2\eta}\big(|B_+|^2 - |B_-|^2\big) \tag{7.3.16}$$

Using the appropriate definitions for $E_{T\pm}$ and η_T, Eq. (7.3.16) can be written in terms of the transverse components for either polarization:

$$\mathcal{P}_z = \frac{1}{2\eta_T}\big(|E_{T+}|^2 - |E_{T-}|^2\big) \tag{7.3.17}$$

As in the normal incidence case, the structure of the matching matrix (7.3.11) implies that (7.3.17) is conserved across the interface.

7.4 *Fresnel Reflection Coefficients*

We look now at the specifics of the Fresnel coefficients (7.3.12) for the two polarization cases. Inserting the two possible definitions (7.2.13) for the transverse refractive indices, we can express ρ_T in terms of the incident and refracted angles:

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{\dfrac{n}{\cos\theta} - \dfrac{n'}{\cos\theta'}}{\dfrac{n}{\cos\theta} + \dfrac{n'}{\cos\theta'}} = \frac{n\cos\theta' - n'\cos\theta}{n\cos\theta' + n'\cos\theta} \\ \rho_{TE} &= \frac{n\cos\theta - n'\cos\theta'}{n\cos\theta + n'\cos\theta'} \end{aligned}} \tag{7.4.1}$$

We note that for normal incidence, $\theta = \theta' = 0$, they both reduce to the usual reflection coefficient $\rho = (n - n')/(n + n')$.[†] Using Snel's law, $n\sin\theta = n'\sin\theta'$, and some trigonometric identities, we may write Eqs. (7.4.1) in a number of equivalent ways. In terms of the angle of incidence only, we have:

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{\sqrt{\left(\dfrac{n'}{n}\right)^2 - \sin^2\theta} - \left(\dfrac{n'}{n}\right)^2\cos\theta}{\sqrt{\left(\dfrac{n'}{n}\right)^2 - \sin^2\theta} + \left(\dfrac{n'}{n}\right)^2\cos\theta} \\ \rho_{TE} &= \frac{\cos\theta - \sqrt{\left(\dfrac{n'}{n}\right)^2 - \sin^2\theta}}{\cos\theta + \sqrt{\left(\dfrac{n'}{n}\right)^2 - \sin^2\theta}} \end{aligned}} \tag{7.4.2}$$

Note that at *grazing angles* of incidence, $\theta \to 90^\circ$, the reflection coefficients tend to $\rho_{TM} \to 1$ and $\rho_{TE} \to -1$, *regardless* of the refractive indices n, n'. One consequence of this property is in wireless communications where the effect of the ground reflections causes the power of the propagating radio wave to attenuate with the *fourth* (instead of the second) power of the distance, thus, limiting the propagation range (see Example 19.3.5.)

We note also that Eqs. (7.4.1) and (7.4.2) remain valid when one or both of the media are lossy. For example, if the right medium is lossy with complex refractive index $n'_c = n'_r - jn'_i$, then, Snel's law, $n\sin\theta = n'_c\sin\theta'$, is still valid but with a complex-valued θ' and (7.4.2) remains the same with the replacement $n' \to n'_c$. The third way of expressing the ρs is in terms of θ, θ' only, without the n, n':

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{\sin 2\theta' - \sin 2\theta}{\sin 2\theta' + \sin 2\theta} = \frac{\tan(\theta' - \theta)}{\tan(\theta' + \theta)} \\ \rho_{TE} &= \frac{\sin(\theta' - \theta)}{\sin(\theta' + \theta)} \end{aligned}} \tag{7.4.3}$$

Fig. 7.4.1 shows the special case of an air-dielectric interface. If the incident wave is from the air side, then Eq. (7.4.2) gives with $n = 1$, $n' = n_d$, where n_d is the (possibly complex-valued) refractive index of the dielectric:

[†] Some references define ρ_{TM} with the opposite sign. Our convention was chosen because it has the expected limit at normal incidence.

$$\boxed{\rho_{TM}=\frac{\sqrt{n_d^2-\sin^2\theta}-n_d^2\cos\theta}{\sqrt{n_d^2-\sin^2\theta}+n_d^2\cos\theta}\,,\quad \rho_{TE}=\frac{\cos\theta-\sqrt{n_d^2-\sin^2\theta}}{\cos\theta+\sqrt{n_d^2-\sin^2\theta}}}\tag{7.4.4}$$

If the incident wave is from inside the dielectric, then we set $n = n_d$ and $n' = 1$:

$$\boxed{\rho_{TM}=\frac{\sqrt{n_d^{-2}-\sin^2\theta}-n_d^{-2}\cos\theta}{\sqrt{n_d^{-2}-\sin^2\theta}+n_d^{-2}\cos\theta}\,,\quad \rho_{TE}=\frac{\cos\theta-\sqrt{n_d^{-2}-\sin^2\theta}}{\cos\theta+\sqrt{n_d^{-2}-\sin^2\theta}}}\tag{7.4.5}$$

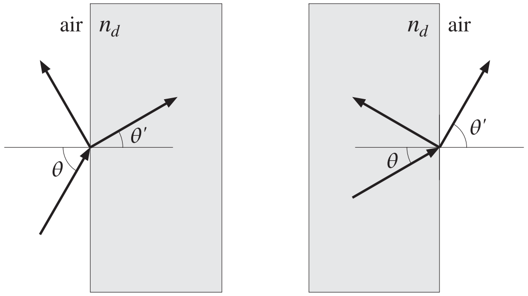


Fig. 7.4.1 Air-dielectric interfaces.

The MATLAB function `fresnel` calculates the expressions (7.4.2) for any range of values of θ. Its usage is as follows:

```
[rtm,rte] = fresnel(na,nb,theta);                % Fresnel reflection coefficients
```

7.5 *Maximum Angle and Critical Angle*

As the incident angle θ varies over $0 \le \theta \le 90^\circ$, the angle of refraction θ' will have a corresponding range of variation. It can be determined by solving for θ' from Snel's law, $n\sin\theta = n'\sin\theta'$:

$$\sin\theta' = \frac{n}{n'}\sin\theta \tag{7.5.1}$$

If $n < n'$ (we assume lossless dielectrics here,) then Eq. (7.5.1) implies that $\sin\theta' = (n/n')\sin\theta < \sin\theta$, or $\theta' < \theta$. Thus, if the incident wave is from a lighter to a denser medium, the refracted angle is always smaller than the incident angle. The maximum value of θ', denoted here by θ'_c, is obtained when θ has its maximum, $\theta = 90^\circ$:

$$\boxed{\sin\theta'_c = \frac{n}{n'}}\qquad\text{(maximum angle of refraction)}\tag{7.5.2}$$

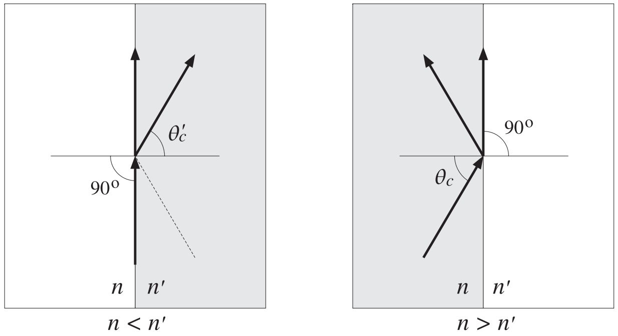


Fig. 7.5.1 Maximum angle of refraction and critical angle of incidence.

Thus, the angle ranges are $0 \le \theta \le 90^\circ$ and $0 \le \theta' \le \theta'_c$. Fig. 7.5.1 depicts this case, as well as the case $n > n'$.

On the other hand, if $n > n'$, and the incident wave is from a denser onto a lighter medium, then $\sin\theta' = (n/n')\sin\theta > \sin\theta$, or $\theta' > \theta$. Therefore, θ' will reach the maximum value of 90° before θ does. The corresponding maximum value of θ satisfies Snel's law, $n\sin\theta_c = n'\sin(\pi/2) = n'$, or,

$$\boxed{\sin\theta_c = \frac{n'}{n}}\qquad\text{(critical angle of incidence)}\tag{7.5.3}$$

This angle is called the *critical angle of incidence.* If the incident wave were from the right, θ_c would be the maximum angle of refraction according to the above discussion.

If $\theta \le \theta_c$, there is normal refraction into the lighter medium. But, if θ exceeds θ_c, the incident wave cannot be refracted and gets completely reflected back into the denser medium. This phenomenon is called *total internal reflection.* Because $n'/n = \sin\theta_c$, we may rewrite the reflection coefficients (7.4.2) in the form:

$$\rho_{TM}=\frac{\sqrt{\sin^2\theta_c-\sin^2\theta}-\sin^2\theta_c\cos\theta}{\sqrt{\sin^2\theta_c-\sin^2\theta}+\sin^2\theta_c\cos\theta}\,,\quad \rho_{TE}=\frac{\cos\theta-\sqrt{\sin^2\theta_c-\sin^2\theta}}{\cos\theta+\sqrt{\sin^2\theta_c-\sin^2\theta}}$$

When $\theta < \theta_c$, the reflection coefficients are real-valued. At $\theta = \theta_c$, they have the values, $\rho_{TM} = -1$ and $\rho_{TE} = 1$. And, when $\theta > \theta_c$, they become complex-valued with unit magnitude. Indeed, switching the sign under the square roots, we have in this case:

$$\rho_{TM}=\frac{-j\sqrt{\sin^2\theta-\sin^2\theta_c}-\sin^2\theta_c\cos\theta}{-j\sqrt{\sin^2\theta-\sin^2\theta_c}+\sin^2\theta_c\cos\theta}\,,\quad \rho_{TE}=\frac{\cos\theta+j\sqrt{\sin^2\theta-\sin^2\theta_c}}{\cos\theta-j\sqrt{\sin^2\theta-\sin^2\theta_c}}$$

where we used the evanescent definition of the square root as discussed in Eqs. (7.7.9) and (7.7.10), that is, we made the replacement

$$\sqrt{\sin^2\theta_c-\sin^2\theta}\;\longrightarrow\;-j\sqrt{\sin^2\theta-\sin^2\theta_c}\,,\quad\text{for }\;\theta\ge\theta_c$$

Both expressions for ρ_T are the ratios of a complex number and its conjugate, and therefore, they are unimodular, $|\rho_{TM}| = |\rho_{TE}| = 1$, for all values of $\theta > \theta_c$. The interface becomes a perfect mirror, with zero transmittance into the lighter medium.

When $\theta > \theta_c$, the fields on the right side of the interface are not zero, but do not propagate away to the right. Instead, they decay exponentially with the distance z. There is no transfer of power (on the average) to the right. To understand this behavior of the fields, we consider the solutions given in Eqs. (7.2.18) and (7.2.20), with no incident field from the right, that is, with $A'_- = B'_- = 0$.

The longitudinal wavenumber in the right medium, k'_z, can be expressed in terms of the angle of incidence θ as follows. We have from Eq. (7.1.7):

$$k_z^2 + k_x^2 = k^2 = n^2k_0^2$$

$$k_z'^2 + k_x'^2 = k'^2 = n'^2k_0^2$$

Because, $k'_x = k_x = k\sin\theta = nk_0\sin\theta$, we may solve for k'_z to get:

$$k_z'^2 = n'^2k_0^2 - k_x'^2 = n'^2k_0^2 - k_x^2 = n'^2k_0^2 - n^2k_0^2\sin^2\theta = k_0^2(n'^2 - n^2\sin^2\theta)$$

or, replacing $n' = n\sin\theta_c$, we find:

$$\boxed{k_z'^2 = n^2k_0^2(\sin^2\theta_c - \sin^2\theta)} \tag{7.5.4}$$

If $\theta \le \theta_c$, the wavenumber k'_z is real-valued and corresponds to ordinary propagating fields that represent the refracted wave. But if $\theta > \theta_c$, we have $k_z'^2 < 0$ and k'_z becomes pure imaginary, say $k'_z = -j\alpha'_z$. The z-dependence of the fields on the right of the interface will be:

$$e^{-jk'_z z} = e^{-\alpha'_z z}\,, \qquad \alpha'_z = nk_0\sqrt{\sin^2\theta - \sin^2\theta_c}$$

Such exponentially decaying fields are called *evanescent waves* because they are effectively confined to within a few multiples of the distance $z = 1/\alpha'_z$ (the penetration length) from the interface.

The maximum value of α'_z, or equivalently, the smallest penetration length $1/\alpha'_z$, is achieved when $\theta = 90^\circ$, resulting in:

$$\alpha'_{\max} = nk_0\sqrt{1 - \sin^2\theta_c} = nk_0\cos\theta_c = k_0\sqrt{n^2 - n'^2}$$

Inspecting Eqs. (7.2.20), we note that the factor $\cos\theta'$ becomes pure imaginary because $\cos^2\theta' = 1 - \sin^2\theta' = 1 - (n/n')^2\sin^2\theta = 1 - \sin^2\theta/\sin^2\theta_c \le 0$, for $\theta \ge \theta_c$. Therefore for either the TE or TM case, the transverse components E_T and H_T will have a 90° phase difference, which will make the time-average power flow into the right medium zero: $\mathcal{P}_z = \mathrm{Re}(E_T H_T^*)/2 = 0$.

Example 7.5.1: Determine the maximum angle of refraction and critical angle of reflection for (a) an air-glass interface and (b) an air-water interface. The refractive indices of glass and water at optical frequencies are: $n_{\text{glass}} = 1.5$ and $n_{\text{water}} = 1.333$.

Solution: There is really only one angle to determine, because if $n = 1$ and $n' = n_{\text{glass}}$, then $\sin(\theta'_c) = n/n' = 1/n_{\text{glass}}$, and if $n = n_{\text{glass}}$ and $n' = 1$, then, $\sin(\theta_c) = n'/n = 1/n_{\text{glass}}$. Thus, $\theta'_c = \theta_c$:

$$\theta_c = \mathrm{asin}\left(\frac{1}{1.5}\right) = 41.8^\circ$$

For the air-water case, we have:

$$\theta_c = \mathrm{asin}\left(\frac{1}{1.333}\right) = 48.6^\circ$$

The refractive index of water at radio frequencies and below is $n_{\text{water}} = 9$ approximately. The corresponding critical angle is $\theta_c = 6.4^\circ$. □

Example 7.5.2: *Prisms.* Glass prisms with 45° angles are widely used in optical instrumentation for bending light beams without the use of metallic mirrors. Fig. 7.5.2 shows two examples.

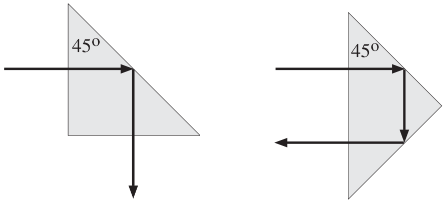


Fig. 7.5.2 Prisms using total internal reflection.

In both cases, the incident beam hits an internal prism side at an angle of 45°, which is greater than the air-glass critical angle of 41.8°. Thus, total internal reflection takes place and the prism side acts as a perfect mirror. □

Example 7.5.3: *Optical Manhole.* Because the air-water interface has $\theta_c = 48.6^\circ$, if we were to view a water surface from above the water, we could only see inside the water within the cone defined by the maximum angle of refraction.

Conversely, were we to view the surface of the water from underneath, we would see the air side only within the critical angle cone, as shown in Fig. 7.5.3. The angle subtended by this cone is 2×48.6 = 97.2°.

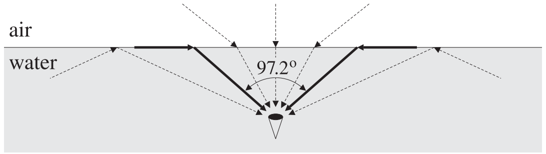


Fig. 7.5.3 Underwater view of the outside world.

The rays arriving from below the surface at an angle greater than θ_c get totally reflected. But because they are weak, the body of water outside the critical cone will appear dark. The critical cone is known as the “optical manhole” [49]. □

Example 7.5.4: *Apparent Depth.* Underwater objects viewed from the outside appear to be closer to the surface than they really are. The apparent depth of the object depends on our viewing angle. Fig. 7.5.4 shows the geometry of the incident and refracted rays.

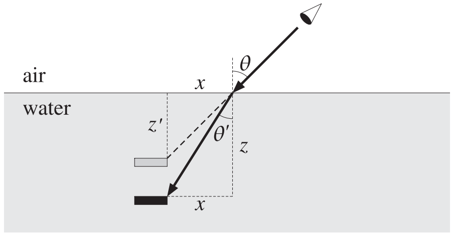


Fig. 7.5.4 Apparent depth of underwater object.

Let θ be the viewing angle and let z and z' be the actual and apparent depths. Our perceived depth corresponds to the extension of the incident ray at angle θ. From the figure, we have: $z = x\cot\theta'$ and $z' = x\cot\theta$. It follows that:

$$z' = \frac{\cot\theta}{\cot\theta'}\,z = \frac{\sin\theta'\cos\theta}{\sin\theta\cos\theta'}\,z$$

Using Snel's law $\sin\theta/\sin\theta' = n'/n = n_{\text{water}}$, we eventually find:

$$z' = \frac{\cos\theta}{\sqrt{n_{\text{water}}^2 - \sin^2\theta}}\,z$$

At normal incidence, we have $z' = z/n_{\text{water}} = z/1.333 = 0.75z$.

Reflection and refraction phenomena are very common in nature. They are responsible for the twinkling and aberration of stars, the flattening of the setting sun and moon, mirages, rainbows, and countless other natural phenomena. Four wonderful expositions of such effects are in Refs. [49–52]. See also the web page [1303]. □

Example 7.5.5: *Optical Fibers.* Total internal reflection is the mechanism by which light is guided along an optical fiber. Fig. 7.5.5 shows a step-index fiber with refractive index n_f surrounded by cladding material of index $n_c < n_f$.

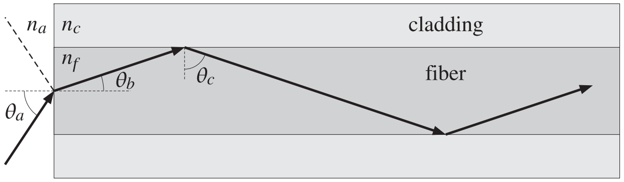


Fig. 7.5.5 Launching a beam into an optical fiber.

If the angle of incidence on the fiber-cladding interface is greater than the critical angle, then total internal reflection will take place. The figure shows a beam launched into the fiber from the air side. The maximum angle of incidence θ_a must be made to correspond to the critical angle θ_c of the fiber-cladding interface. Using Snel's laws at the two interfaces, we have:

$$\sin\theta_a = \frac{n_f}{n_a}\sin\theta_b\,, \quad \sin\theta_c = \frac{n_c}{n_f}$$

Noting that $\theta_b = 90^{\text{o}} - \theta_c$, we find:

$$\sin\theta_a = \frac{n_f}{n_a}\cos\theta_c = \frac{n_f}{n_a}\sqrt{1-\sin^2\theta_c} = \frac{\sqrt{n_f^2 - n_c^2}}{n_a}$$

For example, with $n_a = 1$, $n_f = 1.49$, and $n_c = 1.48$, we find $\theta_c = 83.4^{\text{o}}$ and $\theta_a = 9.9^{\text{o}}$. The angle θ_a is called the *acceptance angle*, and the quantity $NA = \sqrt{n_f^2 - n_c^2}$, the *numerical aperture* of the fiber. □

Example 7.5.6: *Fresnel Rhomb.* The Fresnel rhomb is a glass prism depicted in Fig. 7.5.6 that acts as a 90° retarder. It converts linear polarization into circular. Its advantage over the birefringent retarders discussed in Sec. 4.1 is that it is frequency-independent or achromatic.

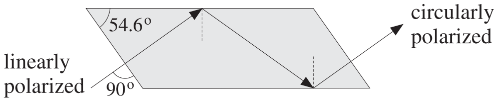


Fig. 7.5.6 Fresnel rhomb.

Assuming a refractive index $n = 1.51$, the critical angle is $\theta_c = 41.47^{\text{o}}$. The angle of the rhomb, $\theta = 54.6^{\text{o}}$, is also the angle of incidence on the internal side. This angle has been chosen such that, at each total internal reflection, the relative phase between the TE and TM polarizations changes by 45°, so that after two reflections it changes by 90°.

The angle of the rhomb can be determined as follows. For $\theta \geq \theta_c$, the reflection coefficients can be written as the unimodular complex numbers:

$$\rho_{TE} = \frac{1+jx}{1-jx}\,, \qquad \rho_{TM} = -\frac{1+jxn^2}{1-jxn^2}\,, \qquad x = \frac{\sqrt{\sin^2\theta - \sin^2\theta_c}}{\cos\theta} \tag{7.5.5}$$

where $\sin\theta_c = 1/n$. It follows that:

$$\rho_{TE} = e^{2j\psi_{TE}}\,, \quad \rho_{TM} = e^{j\pi + 2j\psi_{TM}}$$

where ψ_{TE}, ψ_{TM} are the phase angles of the numerators, that is,

$$\tan\psi_{TE} = x\,, \quad \tan\psi_{TM} = xn^2$$

The relative phase change between the TE and TM polarizations will be:

$$\frac{\rho_{TM}}{\rho_{TE}} = e^{2j\psi_{TM} - 2j\psi_{TE} + j\pi}$$

It is enough to require that $\psi_{TM}-\psi_{TE}=\pi/8$ because then, after two reflections, we will have a 90° change:

$$\frac{\rho_{TM}}{\rho_{TE}} = e^{j\pi/4+j\pi} \quad\Rightarrow\quad \left(\frac{\rho_{TM}}{\rho_{TE}}\right)^2 = e^{j\pi/2+2j\pi} = e^{j\pi/2}$$

From the design condition $\psi_{TM}-\psi_{TE}=\pi/8$, we obtain the required value of x and then of θ. Using a trigonometric identity, we have:

$$\tan(\psi_{TM}-\psi_{TE}) = \frac{\tan\psi_{TM}-\tan\psi_{TE}}{1+\tan\psi_{TM}\tan\psi_{TE}} = \frac{xn^2-x}{1+n^2x^2} = \tan\big(\frac{\pi}{8}\big)$$

This gives the quadratic equation for x:

$$x^2-\frac{1}{\tan(\pi/8)}\big(1-\frac{1}{n^2}\big)\,x+\frac{1}{n^2} = x^2-\frac{\cos^2\theta_c}{\tan(\pi/8)}\,x+\sin^2\theta_c = 0 \tag{7.5.6}$$

Inserting the two solutions of (7.5.6) into Eq. (7.5.5), we may solve for $\sin\theta$, obtaining two possible solutions for θ:

$$\sin\theta = \sqrt{\frac{x^2+\sin^2\theta_c}{x^2+1}} \tag{7.5.7}$$

We may also eliminate x and express the design condition directly in terms of θ:

$$\frac{\cos\theta\sqrt{\sin^2\theta-\sin^2\theta_c}}{\sin^2\theta} = \tan\big(\frac{\pi}{8}\big) \tag{7.5.8}$$

However, the two-step process is computationally more convenient. For $n=1.51$, we find the two roots of Eq. (7.5.6): $x=0.822$ and $x=0.534$. Then, (7.5.7) gives the two values $\theta=54.623°$ and $\theta=48.624°$. The rhomb could just as easily be designed with the second value of θ.

For $n=1.50$, we find the angles $\theta=53.258°$ and $50.229°$. For $n=1.52$, we have $\theta=55.458°$ and $47.553°$. See Problem 7.5 for an equivalent approach. □

Example 7.5.7: *Goos-Hänchen Effect.* When a beam of light is reflected obliquely from a denser-to-rarer interface at an angle greater than the TIR angle, it suffers a lateral displacement, relative to the ordinary reflected ray, known as the Goos-Hänchen shift, as shown Fig. 7.5.7.

Let n, n' be the refractive indices of the two media with $n>n'$, and consider first the case of ordinary reflection at an incident angle $\theta_0<\theta_c$. For a plane wave with a free-space wavenumber $k_0=\omega/c_0$ and wavenumber components $k_x=k_0n\sin\theta_0$, $k_z=k_0n\cos\theta_0$, the corresponding incident, reflected, and transmitted transverse electric fields will be:

$$\begin{aligned}
E_i(x,z) &= e^{-jk_xx}e^{-jk_zz}\\
E_r(x,z) &= \rho(k_x)\,e^{-jk_xx}e^{+jk_zz}\\
E_t(x,z) &= \tau(k_x)\,e^{-jk_xx}e^{-jk'_zz}\,,\quad k'_z=\sqrt{k_0^2n'^2-k_x^2}
\end{aligned}$$

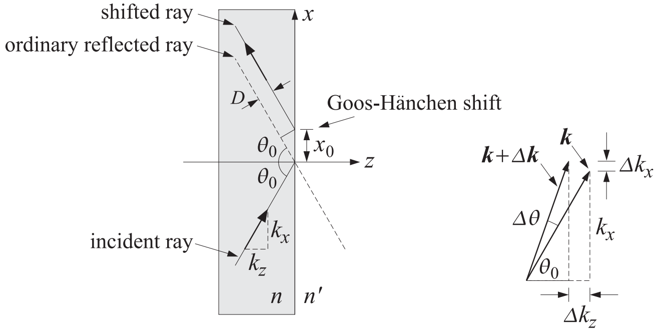


Fig. 7.5.7 Goos-Hänchen shift, with $n_a>n_b$ and $\theta_0>\theta_c$.

where $\rho(k_x)$ and $\tau(k_x)=1+\rho(k_x)$ are the transverse reflection and transmission coefficients, viewed as functions of k_x. For TE and TM polarizations, $\rho(k_x)$ is given by

$$\rho_{TE}(k_x) = \frac{k_z-k'_z}{k_z+k'_z}\,,\quad \rho_{TM}(k_x) = \frac{k'_zn^2-k_zn'^2}{k'_zn^2+k_zn'^2}$$

A beam can be made up by forming a linear combination of such plane waves having a small spread of angles about θ_0. For example, consider a second plane wave with wavenumber components $k_x+\Delta k_x$ and $k_z+\Delta k_z$. These must satisfy $(k_x+\Delta k_x)^2+(k_z+\Delta k_z)^2 = k_x^2+k_z^2=k_0^2n^2$, or to lowest order in Δk_x,

$$k_x\Delta k_x+k_z\Delta k_z = 0 \quad\Rightarrow\quad \Delta k_z = -\Delta k_x\frac{k_x}{k_z} = -\Delta k_x\tan\theta_0$$

Similarly, we have for the transmitted wavenumber $\Delta k'_z=-\Delta k_x\tan\theta'_0$, where θ'_0 is given by Snel's law, $n\sin\theta_0=n'\sin\theta'_0$. The incident, reflected, and transmitted fields will be given by the sum of the two plane waves:

$$\begin{aligned}
E_i(x,z) &= e^{-jk_xx}e^{-jk_zz}+e^{-j(k_x+\Delta k_x)x}e^{-j(k_z+\Delta k_z)z}\\
E_r(x,z) &= \rho(k_x)\,e^{-jk_xx}e^{+jk_zz}+\rho(k_x+\Delta k_x)\,e^{-j(k_x+\Delta k_x)x}e^{+j(k_z+\Delta k_z)z}\\
E_t(x,z) &= \tau(k_x)\,e^{-jk_xx}e^{-jk'_zz}+\tau(k_x+\Delta k_x)\,e^{-j(k_x+\Delta k_x)x}e^{-j(k'_z+\Delta k'_z)z}
\end{aligned}$$

Replacing $\Delta k_z=-\Delta k_x\tan\theta_0$ and $\Delta k'_z=-\Delta k_x\tan\theta'_0$, we obtain:

$$\begin{aligned}
E_i(x,z) &= e^{-jk_xx}e^{-jk_zz}\big[1+e^{-j\Delta k_x(x-z\tan\theta_0)}\big]\\
E_r(x,z) &= e^{-jk_xx}e^{+jk_zz}\big[\rho(k_x)+\rho(k_x+\Delta k_x)\,e^{-j\Delta k_x(x+z\tan\theta_0)}\big]\\
E_t(x,z) &= e^{-jk_xx}e^{-jk'_zz}\big[\tau(k_x)+\tau(k_x+\Delta k_x)\,e^{-j\Delta k_x(x-z\tan\theta'_0)}\big]
\end{aligned} \tag{7.5.9}$$

The incidence angle of the second wave is $\theta_0+\Delta\theta$, where $\Delta\theta$ is obtained by expanding $k_x+\Delta k_x=k_0n\sin(\theta_0+\Delta\theta)$ to first order, or, $\Delta k_x=k_0n\cos\theta_0\,\Delta\theta$. If we assume that $\theta_0<\theta_c$, as well as $\theta_0+\Delta\theta<\theta_c$, then $\rho(k_x)$ and $\rho(k_x+\Delta k_x)$ are both real-valued. It follows that the two terms in the reflected wave $E_r(x,z)$ will differ by a small amplitude change and therefore we can set $\rho(k_x + \Delta k_x) \simeq \rho(k_x)$. Similarly, in the transmitted field we may set $\tau(k_x + \Delta k_x) \simeq \tau(k_x)$. Thus, when $\theta_0 < \theta_c$, Eq. (7.5.9) reads approximately

$$\begin{aligned} E_i(x,z) &= e^{-jk_x x} e^{-jk_z z}\big[1 + e^{-j\Delta k_x (x - z\tan\theta_0)}\big] \\ E_r(x,z) &= \rho(k_x) e^{-jk_x x} e^{+jk_z z}\big[1 + e^{-j\Delta k_x (x + z\tan\theta_0)}\big] \\ E_t(x,z) &= \tau(k_x) e^{-jk_x x} e^{-jk'_z z}\big[1 + e^{-j\Delta k_x (x - z\tan\theta'_0)}\big] \end{aligned} \qquad (7.5.10)$$

Noting that $\big|1 + e^{-j\Delta k_x (x - z\tan\theta_0)}\big| \leq 2$, with equality achieved when $x - z\tan\theta_0 = 0$, it follows that the intensities of these waves are maximized along the ordinary geometric rays defined by the beam angles θ_0 and θ'_0, that is, along the straight lines:

$$\begin{aligned} x - z\tan\theta_0 &= 0\,, \quad \text{incident ray} \\ x + z\tan\theta_0 &= 0\,, \quad \text{reflected ray} \\ x - z\tan\theta'_0 &= 0\,, \quad \text{transmitted ray} \end{aligned} \qquad (7.5.11)$$

On the other hand, if $\theta_0 > \theta_c$ and $\theta_0 + \Delta\theta > \theta_c$, the reflection coefficients become unimodular complex numbers, as in Eq. (7.5.5). Writing $\rho(k_x) = e^{j\phi(k_x)}$, Eq. (7.5.9) gives:

$$E_r(x,z) = e^{-jk_x x} e^{+jk_z z}\big[e^{j\phi(k_x)} + e^{j\phi(k_x + \Delta k_x)} e^{-j\Delta k_x (x + z\tan\theta_0)}\big] \qquad (7.5.12)$$

Introducing the Taylor series expansion, $\phi(k_x + \Delta k_x) \simeq \phi(k_x) + \Delta k_x\, \phi'(k_x)$, we obtain:

$$E_r(x,z) = e^{j\phi(k_x)} e^{-jk_x x} e^{+jk_z z}\big[1 + e^{j\Delta k_x \phi'(k_x)} e^{-j\Delta k_x (x + z\tan\theta_0)}\big]$$

Setting $x_0 = \phi'(k_x)$, we have:

$$E_r(x,z) = e^{j\phi(k_x)} e^{-jk_x x} e^{+jk_z z}\big[1 + e^{-j\Delta k_x (x - x_0 + z\tan\theta_0)}\big] \qquad (7.5.13)$$

This implies that the maximum intensity of the reflected beam will now be along the shifted ray defined by:

$$x - x_0 + z\tan\theta_0 = 0\,, \quad \text{shifted reflected ray} \qquad (7.5.14)$$

Thus, the origin of the Goos-Hänchen shift can be traced to the relative phase shifts arising from the reflection coefficients in the plane-wave components making up the beam. The parallel displacement, denoted by D in Fig. 7.5.7, is related to x_0 by $D = x_0 \cos\theta_0$. Noting that $dk_x = k_0 n \cos\theta\, d\theta$, we obtain

$$\boxed{D = \cos\theta_0 \frac{d\phi}{dk_x} = \frac{1}{k_0 n} \frac{d\phi}{d\theta}\bigg|_{\theta_0}} \quad \text{(Goos-Hänchen shift)} \qquad (7.5.15)$$

Using Eq. (7.5.5), we obtain the shifts for the TE and TM cases:

$$D_{TE} = \frac{2\sin\theta_0}{k_0 n \sqrt{\sin^2\theta_0 - \sin^2\theta_c}}\,, \quad D_{TM} = \frac{n'^2 D_{TE}}{(n^2+1)\sin^2\theta_0 - n'^2} \qquad (7.5.16)$$

These expressions are not valid near the critical angle $\theta_0 \simeq \theta_c$ because then the Taylor series expansion for $\phi(k_x)$ cannot be justified. □

Besides its its use in optical fibers, total internal reflection has several other applications [516–552], such as internal reflection spectroscopy, chemical and biological sensors, fingerprint identification, surface plasmon resonance, and high resolution microscopy.

7.6 *Brewster Angle*

The Brewster angle is that angle of incidence at which the TM Fresnel reflection coefficient vanishes, $\rho_{TM} = 0$. The TE coefficient ρ_{TE} cannot vanish for any angle θ, for non-magnetic materials. A scattering model of Brewster's law is discussed in [653]. Fig. 7.6.1 depicts the Brewster angles from either side of an interface.

The Brewster angle is also called the *polarizing* angle because if a mixture of TM and TE waves are incident on a dielectric interface at that angle, only the TE or perpendicularly polarized waves will be reflected. This is not necessarily a good method of generating polarized waves because even though ρ_{TE} is non-zero, it may be too small to provide a useful amount of reflected power. Better polarization methods are based on using (a) multilayer structures with alternating low/high refractive indices and (b) birefringent and dichroic materials, such as calcite and polaroids.

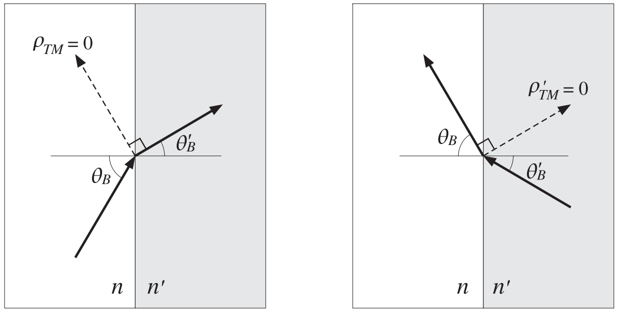


Fig. 7.6.1 Brewster angles.

The Brewster angle θ_B is determined by the condition, $\rho_{TM} = 0$, in Eq. (7.4.2). Setting the numerator of that expression to zero, we have:

$$\sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\theta_B} = \left(\frac{n'}{n}\right)^2 \cos\theta_B \qquad (7.6.1)$$

After some algebra, we obtain the alternative expressions:

$$\boxed{\sin\theta_B = \frac{n'}{\sqrt{n^2 + n'^2}}} \quad \Leftrightarrow \quad \boxed{\tan\theta_B = \frac{n'}{n}} \quad \text{(Brewster angle)} \qquad (7.6.2)$$

Similarly, the Brewster angle θ'_B from the other side of the interface is:

$$\boxed{\sin\theta'_B = \frac{n}{\sqrt{n^2 + n'^2}}} \quad \Leftrightarrow \quad \boxed{\tan\theta'_B = \frac{n}{n'}} \quad \text{(Brewster angle)} \qquad (7.6.3)$$

The angle θ'_B is related to θ_B by Snel's law, $n' \sin\theta'_B = n \sin\theta_B$, and corresponds to zero reflection from that side, $\rho'_{TM} = -\rho_{TM} = 0$. A consequence of Eq. (7.6.2) is that $\theta_B = 90^{\circ} - \theta'_B$, or, $\theta_B + \theta'_B = 90^{\circ}$. Indeed,

$$\frac{\sin\theta_B}{\cos\theta_B} = \tan\theta_B = \frac{n'}{n} = \frac{\sin\theta_B}{\sin\theta'_B}$$

which implies $\cos\theta_B = \sin\theta'_B$, or $\theta_B = 90^{\text{o}} - \theta'_B$. The same conclusion can be reached immediately from Eq. (7.4.3). Because, $\theta'_B - \theta_B \neq 0$, the only way for the ratio of the two tangents to vanish is for the denominator to be infinity, that is, $\tan(\theta'_B + \theta_B) = \infty$, or, $\theta_B + \theta'_B = 90^{\text{o}}$.

As shown in Fig. 7.6.1, the angle of the refracted ray with the would-be reflected ray is 90^{o}. Indeed, this angle is $180^{\text{o}} - (\theta'_B + \theta_B) = 180^{\text{o}} - 90^{\text{o}} = 90^{\text{o}}$.

The TE reflection coefficient at θ_B can be calculated very simply by using Eq. (7.6.1) into (7.4.2). After canceling a common factor of $\cos\theta_B$, we find:

$$\rho_{TE}(\theta_B) = \frac{1 - \left(\dfrac{n'}{n}\right)^2}{1 + \left(\dfrac{n'}{n}\right)^2} = \frac{n^2 - n'^2}{n^2 + n'^2} \tag{7.6.4}$$

Example 7.6.1: *Brewster angles for water.* The Brewster angles from the air and the water sides of an air-water interface are:

$$\theta_B = \text{atan}\left(\frac{1.333}{1}\right) = 53.1^{\text{o}}\,, \quad \theta'_B = \text{atan}\left(\frac{1}{1.333}\right) = 36.9^{\text{o}}$$

We note that $\theta_B + \theta'_B = 90^{\text{o}}$. At RF, the refractive index is $n_{\text{water}} = 9$ and we find $\theta_B = 83.7^{\text{o}}$ and $\theta'_B = 6.3^{\text{o}}$. We also find $\rho_{TE}(\theta_B) = -0.2798$ and $|\rho_{TE}(\theta_B)|^2 = 0.0783/$ Thus, for TE waves, only 7.83% of the incident power gets reflected at the Brewster angle. □

Example 7.6.2: *Brewster Angles for Glass.* The Brewster angles for the two sides of an air-glass interface are:

$$\theta_B = \text{atan}\left(\frac{1.5}{1}\right) = 56.3^{\text{o}}\,, \quad \theta'_B = \text{atan}\left(\frac{1}{1.5}\right) = 33.7^{\text{o}}$$

Fig. 7.6.2 shows the reflection coefficients $|\rho_{TM}(\theta)|, |\rho_{TE}(\theta)|$ as functions of the angle of incidence θ from the air side, calculated with the MATLAB function `fresnel`.

Both coefficients start at their normal-incidence value $|\rho| = |(1-1.5)/(1+1.5)| = 0.2$ and tend to unity at grazing angle $\theta = 90^{\text{o}}$. The TM coefficient vanishes at the Brewster angle $\theta_B = 56.3^{\text{o}}$.

The right graph in the figure depicts the reflection coefficients $|\rho'_{TM}(\theta')|, |\rho'_{TE}(\theta')|$ as functions of the incidence angle θ' from the glass side. Again, the TM coefficient vanishes at the Brewster angle $\theta'_B = 33.7^{\text{o}}$. The typical MATLAB code for generating this graph was:

```
na = 1; nb = 1.5;
[thb,thc]  = brewster(na,nb);           % calculate Brewster angle
th = linspace(0,90,901);                % equally-spaced angles at 0.1° intervals
[rte,rtm] = fresnel(na,nb,th);          % Fresnel reflection coefficients
plot(th,abs(rtm), th,abs(rte));
```

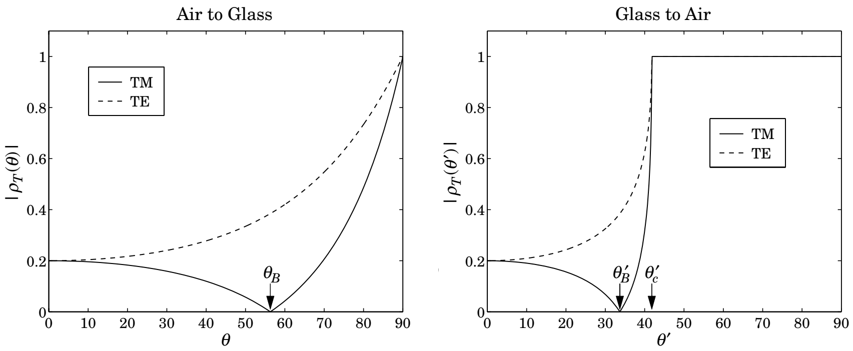


Fig. 7.6.2 TM and TE reflection coefficients versus angle of incidence.

The critical angle of reflection is in this case $\theta'_c = \text{asin}(1/1.5) = 41.8^{\text{o}}$. As soon as θ' exceeds θ'_c, both coefficients become complex-valued with unit magnitude.

The value of the TE reflection coefficient at the Brewster angle is $\rho_{TE} = -\rho'_{TE} = -0.38$, and the TE reflectance $|\rho_{TE}|^2 = 0.144$, or 14.4 percent. This is too small to be useful for generating TE polarized waves by reflection.

Two properties are evident from Fig. 7.6.2. One is that $|\rho_{TM}| \leq |\rho_{TE}|$ for all angles of incidence. The other is that $\theta'_B \leq \theta'_c$. Both properties can be proved in general. □

Example 7.6.3: *Lossy dielectrics.* The Brewster angle loses its meaning if one of the media is lossy. For example, assuming a complex refractive index for the dielectric, $n_d = n_r - jn_i$, we may still calculate the reflection coefficients from Eq. (7.4.4). It follows from Eq. (7.6.2) that the Brewster angle θ_B will be complex-valued.

Fig. 7.6.3 shows the TE and TM reflection coefficients versus the angle of incidence θ (from air) for the two cases $n_d = 1.50 - 0.15j$ and $n_d = 1.50 - 0.30j$ and compares them with the lossless case of $n_d = 1.5$. (The values for n_i were chosen only for plotting purposes and have no physical significance.)

The curves retain much of their lossless shape, with the TM coefficient having a minimum near the lossless Brewster angle. The larger the extinction coefficient n_i, the larger the deviation from the lossless case. In the next section, we discuss reflection from lossy media in more detail. □

7.7 *Complex Waves*

In this section, we discuss some examples of complex waves that appear in oblique incidence problems. We consider the cases of (a) total internal reflection, (b) reflection from and refraction into a lossy medium, (c) the Zenneck surface wave, and (d) surface plasmons. Further details may be found in [869–876] and [1116].

Because the wave numbers become complex-valued, e.g., $\boldsymbol{k} = \boldsymbol{\beta} - j\boldsymbol{\alpha}$, the angle of refraction and possibly the angle of incidence may become complex-valued. To avoid

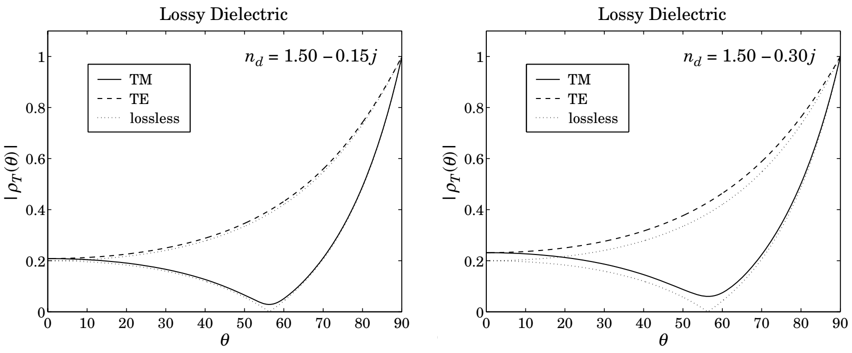


Fig. 7.6.3 TM and TE reflection coefficients for lossy dielectric.

unnecessary complex algebra, it proves convenient to recast impedances, reflection coefficients, and field expressions in terms of wavenumbers. This can be accomplished by making substitutions such as $\cos\theta = k_z/k$ and $\sin\theta = k_x/k$.

Using the relationships $k\eta = \omega\mu$ and $k/\eta = \omega\epsilon$, we may rewrite the TE and TM transverse impedances in the forms:

$$\boxed{\eta_{TE} = \frac{\eta}{\cos\theta} = \frac{\eta k}{k_z} = \frac{\omega\mu}{k_z}\,, \qquad \eta_{TM} = \eta\cos\theta = \frac{\eta k_z}{k} = \frac{k_z}{\omega\epsilon}} \tag{7.7.1}$$

We consider an interface geometry as shown in Fig. 7.1.1 and assume that there are no incident fields from the right of the interface. Snel's law implies that $k_x = k'_x$, where $k_x = k\sin\theta = \omega\sqrt{\mu_0\epsilon}\,\sin\theta$, if the incident angle is real-valued.

Assuming non-magnetic media from both sides of an interface ($\mu = \mu' = \mu_0$), the TE and TM transverse reflection coefficients will take the forms:

$$\boxed{\rho_{TE} = \frac{\eta'_{TE} - \eta_{TE}}{\eta'_{TE} + \eta_{TE}} = \frac{k_z - k'_z}{k_z + k'_z}\,, \qquad \rho_{TM} = \frac{\eta'_{TM} - \eta_{TM}}{\eta'_{TM} + \eta_{TM}} = \frac{k'_z\epsilon - k_z\epsilon'}{k'_z\epsilon + k_z\epsilon'}} \tag{7.7.2}$$

The corresponding transmission coefficients will be:

$$\tau_{TE} = 1 + \rho_{TE} = \frac{2k_z}{k_z + k'_z}\,, \qquad \tau_{TM} = 1 + \rho_{TM} = \frac{2k'_z\epsilon}{k'_z\epsilon + k_z\epsilon'} \tag{7.7.3}$$

We can now rewrite Eqs. (7.2.18) and (7.2.20) in terms of transverse amplitudes and transverse reflection and transmission coefficients. Defining $E_0 = A_+\cos\theta$ or $E_0 = B_+$ in the TM or TE cases and replacing $\tan\theta = k_x/k_z$, $\tan\theta' = k'_x/k'_z = k_x/k'_z$, we have for

the TE case for the fields at the left and right sides of the interface:

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= \hat{\mathbf{y}}\,E_0\big[e^{-jk_z z} + \rho_{TE}\,e^{jk_z z}\big]e^{-jk_x x}\\
\boldsymbol{H}(\mathbf{r}) &= \frac{E_0}{\eta_{TE}}\left[\left(-\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z z} + \rho_{TE}\left(\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z z}\right]e^{-jk_x x}\\
\boldsymbol{E}'(\mathbf{r}) &= \hat{\mathbf{y}}\,\tau_{TE}\,E_0 e^{-jk'_z z}e^{-jk_x x}\\
\boldsymbol{H}'(\mathbf{r}) &= \frac{\tau_{TE}\,E_0}{\eta'_{TE}}\left(-\hat{\mathbf{x}} + \frac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{-jk'_z z}e^{-jk_x x}
\end{aligned}} \quad \text{(TE)} \tag{7.7.4}$$

and for the TM case:

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= E_0\left[\left(\hat{\mathbf{x}} - \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z z} + \rho_{TM}\left(\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z z}\right]e^{-jk_x x}\\
\boldsymbol{H}(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{E_0}{\eta_{TM}}\big[e^{-jk_z z} - \rho_{TM}\,e^{jk_z z}\big]e^{-jk_x x}\\
\boldsymbol{E}'(\mathbf{r}) &= \tau_{TM}\,E_0\left(\hat{\mathbf{x}} - \frac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{-jk'_z z}e^{-jk_x x}\\
\boldsymbol{H}'(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{\tau_{TM}\,E_0}{\eta'_{TM}}e^{-jk'_z z}e^{-jk_x x}
\end{aligned}} \quad \text{(TM)} \tag{7.7.5}$$

Equations (7.7.4) and (7.7.5) are dual to each other, as are Eqs. (7.7.1). They transform into each other under the duality transformation $\boldsymbol{E} \to \boldsymbol{H}$, $\boldsymbol{H} \to -\boldsymbol{E}$, $\epsilon \to \mu$, and $\mu \to \epsilon$. See Sec. 17.2 for more on the concept of duality.

In all of our complex-wave examples, the transmitted wave will be complex with $\boldsymbol{k}' = k_x\hat{\mathbf{x}} + k'_z\hat{\mathbf{z}} = \boldsymbol{\beta}' - j\boldsymbol{\alpha}' = (\beta_x - j\alpha_x)\hat{\mathbf{x}} + (\beta'_z - j\alpha'_z)\hat{\mathbf{z}}$. This must satisfy the constraint $\boldsymbol{k}'\cdot\boldsymbol{k}' = \omega^2\mu_0\epsilon'$. Thus, the space dependence of the transmitted fields will have the general form:

$$e^{-jk'_z z}e^{-jk_x x} = e^{-j(\beta'_z - j\alpha'_z)z}e^{-j(\beta_x - j\alpha_x)x} = e^{-(\alpha'_z z + \alpha_x x)}\,e^{-j(\beta'_z z + \beta_x x)} \tag{7.7.6}$$

For the wave to attenuate at large distances into the right medium, it is required that $\alpha'_z > 0$. Except for the Zenneck-wave case, which has $\alpha_x > 0$, all other examples will have $\alpha_x = 0$, corresponding to a real-valued wavenumber $k'_x = k_x = \beta_x$. Fig. 7.7.1 shows the constant-amplitude and constant-phase planes within the transmitted medium defined, respectively, by:

$$\alpha'_z z + \alpha_x x = \text{const.}\,, \qquad \beta'_z z + \beta_x x = \text{const.} \tag{7.7.7}$$

As shown in the figure, the corresponding angles ϕ and ψ that the vectors $\boldsymbol{\beta}'$ and $\boldsymbol{\alpha}'$ form with the z-axis are given by:

$$\tan\phi = \frac{\beta_x}{\beta'_z}\,, \qquad \tan\psi = \frac{\alpha_x}{\alpha'_z} \tag{7.7.8}$$

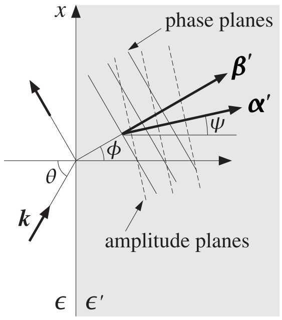


Fig. 7.7.1 Constant-phase and constant-amplitude planes for the transmitted wave.

The wave numbers k_z, k_z' are related to k_x through

$$k_z^2 = \omega^2\mu\epsilon - k_x^2\,,\quad k_z'^2 = \omega^2\mu\epsilon' - k_x^2$$

In calculating k_z and k_z' by taking square roots of the above expressions, it is necessary, in complex-waves problems, to get the correct signs of their imaginary parts, such that evanescent waves are described correctly. This leads us to define an “evanescent” square root as follows. Let $\epsilon = \epsilon_R - j\epsilon_I$ with $\epsilon_I > 0$ for an absorbing medium, then

$$k_z = \text{sqrte}\big(\omega^2\mu(\epsilon_R - j\epsilon_I) - k_x^2\big) = \begin{cases} \sqrt{\omega^2\mu(\epsilon_R - j\epsilon_I) - k_x^2}\,, & \text{if } \epsilon_I \neq 0 \\ -j\sqrt{k_x^2 - \omega^2\mu\epsilon_R}\,, & \text{if } \epsilon_I = 0 \end{cases} \tag{7.7.9}$$

If $\epsilon_I = 0$ and $\omega^2\mu\epsilon_R - k_x^2 > 0$, then the two expressions give the same answer. But if $\epsilon_I = 0$ and $\omega^2\mu\epsilon_R - k_x^2 < 0$, then k_z is correctly calculated from the second expression. The MATLAB function `sqrte.m` implements the above definition. It is defined by

$$y = \text{sqrte}(z) = \begin{cases} -j\sqrt{|z|}\,, & \text{if } \operatorname{Re}(z) < 0 \text{ and } \operatorname{Im}(z) = 0 \\ \sqrt{z}\,, & \text{otherwise} \end{cases} \qquad \text{(evanescent SQRT)} \tag{7.7.10}$$

Some examples of the issues that arise in taking such square roots are elaborated in the next few sections.

7.8 Total Internal Reflection

We already discussed this case in Sec. 7.5. Here, we look at it from the point of view of complex-waves. Both media are assumed to be lossless, but with $\epsilon > \epsilon'$. The angle of incidence θ will be real, so that $k_x' = k_x = k\sin\theta$ and $k_z = k\cos\theta$, with $k = \omega\sqrt{\mu_0\epsilon}$. Setting $k_z' = \beta_z' - j\alpha_z'$, we have the constraint equation:

$$k_x'^2 + k_z'^2 = k'^2 \quad\Rightarrow\quad k_z'^2 = (\beta_z' - j\alpha_z')^2 = \omega^2\mu_0\epsilon' - k_x^2 = \omega^2\mu_0(\epsilon' - \epsilon\sin^2\theta)$$

which separates into the real and imaginary parts:

$$\begin{aligned} \beta_z'^2 - \alpha_z'^2 &= \omega^2\mu_0(\epsilon' - \epsilon\sin^2\theta) = k^2(\sin^2\theta_c - \sin^2\theta) \\ \alpha_z'\beta_z' &= 0 \end{aligned} \tag{7.8.1}$$

where we set $\sin^2\theta_c = \epsilon'/\epsilon$ and $k^2 = \omega^2\mu_0\epsilon$. This has two solutions: (a) $\alpha_z' = 0$ and $\beta_z'^2 = k^2(\sin^2\theta_c - \sin^2\theta)$, valid when $\theta \leq \theta_c$, and (b) $\beta_z' = 0$ and $\alpha_z'^2 = k^2(\sin^2\theta - \sin^2\theta_c)$, valid when $\theta \geq \theta_c$.

Case (a) corresponds to ordinary refraction into the right medium, and case (b), to total internal reflection. In the latter case, we have $k_z' = -j\alpha_z'$ and the TE and TM reflection coefficients (7.7.2) become unimodular complex numbers:

$$\rho_{TE} = \frac{k_z - k_z'}{k_z + k_z'} = \frac{k_z + j\alpha_z'}{k_z - j\alpha_z'}\,, \qquad \rho_{TM} = \frac{k_z'\epsilon - k_z\epsilon'}{k_z'\epsilon + k_z\epsilon'} = -\frac{k_z\epsilon' + j\alpha_z'\epsilon}{k_z\epsilon' - j\alpha_z'\epsilon}$$

The complete expressions for the fields are given by Eqs. (7.7.4) or (7.7.5). The propagation phase factor in the right medium will be in case (b):

$$e^{-jk_z'z}e^{-jk_xx} = e^{-\alpha_z'z}e^{-jk_xx}$$

Thus, the constant-phase planes are the constant-x planes ($\phi = 90^{\text{o}}$), or, the yz-planes. The constant-amplitude planes are the constant-z planes ($\psi = 0^{\text{o}}$), or, the xy-planes, as shown in Fig. 7.8.1.

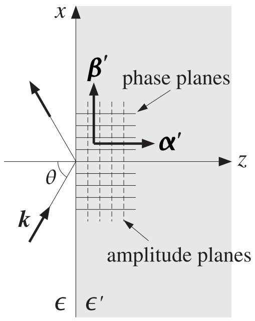


Fig. 7.8.1 Constant-phase and constant-amplitude planes for total internal reflection ($\theta \geq \theta_c$).

7.9 Oblique Incidence on a Lossy Medium

Here, we assume a lossless medium on the left side of the interface and a lossy one, such as a conductor, on the right. The effective dielectric constant ϵ' of the lossy medium is specified by its real and imaginary parts, as in Eq. (2.6.2):

$$\epsilon' = \epsilon_d' - j\left(\epsilon_d'' + \frac{\sigma}{\omega}\right) = \epsilon_R' - j\epsilon_I' \tag{7.9.1}$$

Equivalently, we may characterize the lossy medium by the real and imaginary parts of the wavenumber k', using Eq. (2.6.12):

$$k' = \beta' - j\alpha' = \omega\sqrt{\mu_0\epsilon'} = \omega\sqrt{\mu_0(\epsilon_R' - j\epsilon_I')} \tag{7.9.2}$$

In the left medium, the wavenumber is real with components $k_x = k\sin\theta$, $k_z = k\cos\theta$, with $k = \omega\sqrt{\mu_0\epsilon}$. In the lossy medium, the wavenumber is complex-valued with components $k_x' = k_x$ and $k_z' = \beta_z' - j\alpha_z'$. Using Eq. (7.9.2) in the condition $\mathbf{k}'\cdot\mathbf{k}' = k'^2$, we obtain:

$$k_x'^2 + k_z'^2 = k'^2 \quad\Rightarrow\quad k_x^2 + (\beta_z' - j\alpha_z')^2 = (\beta' - j\alpha')^2 = \omega^2\mu_0(\epsilon_R' - j\epsilon_I') \tag{7.9.3}$$

which separates into its real and imaginary parts:

$$\begin{gathered}\beta_z'^2 - \alpha_z'^2 = \beta'^2 - \alpha'^2 - k_x^2 = \omega^2\mu_0\epsilon_R' - k_x^2 = \omega^2\mu_0(\epsilon_R' - \epsilon\sin^2\theta) \equiv D_R\\ 2\beta_z'\alpha_z' = 2\beta'\alpha' = \omega^2\mu_0\epsilon_I' \equiv D_I\end{gathered} \tag{7.9.4}$$

where we replaced $k_x^2 = k^2\sin^2\theta = \omega^2\mu_0\epsilon\sin^2\theta$. The solutions of Eqs. (7.9.4) leading to a non-negative α_z' are:

$$\beta_z' = \left[\frac{\sqrt{D_R^2 + D_I^2} + D_R}{2}\right]^{1/2}, \quad \alpha_z' = \left[\frac{\sqrt{D_R^2 + D_I^2} - D_R}{2}\right]^{1/2} \tag{7.9.5}$$

For MATLAB implementation, it is simpler to solve Eq. (7.9.3) directly as a complex square root (but see also Eq. (7.9.10)):

$$k_z' = \beta_z' - j\alpha_z' = \sqrt{k'^2 - k_x^2} = \sqrt{\omega^2\mu_0(\epsilon_R' - j\epsilon_I') - k_x^2} = \sqrt{D_R - jD_I} \tag{7.9.6}$$

Eqs. (7.9.5) define completely the reflection coefficients (7.7.2) and the field solutions for both TE and TM waves given by Eqs. (7.7.4) and (7.7.5). Within the lossy medium the transmitted fields will have space-dependence:

$$e^{-jk_z'z}e^{-jk_xx} = e^{-\alpha_z'z}e^{-j(\beta_z'z + k_xx)}$$

The fields attenuate exponentially with distance z. The constant phase and amplitude planes are shown in Fig. 7.9.1.

For the reflected fields, the TE and TM reflection coefficients are given by Eqs. (7.7.2). If the incident wave is linearly polarized having both TE and TM components, the corresponding reflected wave will be elliptically polarized because the ratio ρ_{TM}/ρ_{TE} is now complex-valued. Indeed, using the relationships $k_x^2 + k_z^2 = \omega^2\mu_0\epsilon$ and $k_x^2 + k_z'^2 = \omega^2\mu_0\epsilon'$ in ρ_{TM} of Eq. (7.7.2), it can be shown that (see Problem 7.5):

$$\frac{\rho_{TM}}{\rho_{TE}} = \frac{k_zk_z' - k_x^2}{k_zk_z' + k_x^2} = \frac{k_z' - k\sin\theta\tan\theta}{k_z' + k\sin\theta\tan\theta} = \frac{\beta_z' - j\alpha_z' - k\sin\theta\tan\theta}{\beta_z' - j\alpha_z' + k\sin\theta\tan\theta} \tag{7.9.7}$$

In the case of a lossless medium, $\epsilon' = \epsilon_R'$ and $\epsilon_I' = 0$, Eq. (7.9.5) gives:

$$\beta_z' = \sqrt{\frac{|D_R| + D_R}{2}}, \quad \alpha_z' = \sqrt{\frac{|D_R| - D_R}{2}} \tag{7.9.8}$$

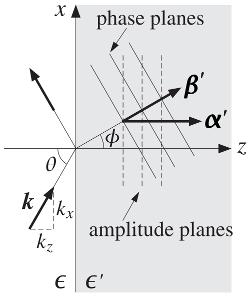


Fig. 7.9.1 Constant-phase and constant-amplitude planes for refracted wave.

If $\epsilon_R' > \epsilon$, then $D_R = \omega^2\mu_0(\epsilon_R' - \epsilon\sin^2\theta)$ is positive for all angles θ, and (7.9.8) gives the expected result $\beta_z' = \sqrt{D_R} = \omega\sqrt{\mu_0(\epsilon_R' - \epsilon\sin^2\theta)}$ and $\alpha_z' = 0$.

On the other hand, in the case of total internal reflection, that is, when $\epsilon_R' < \epsilon$, the quantity D_R is positive for angles $\theta < \theta_c$, and negative for $\theta > \theta_c$, where the critical angle is defined through $\epsilon_R' = \epsilon\sin^2\theta_c$ so that $D_R = \omega^2\mu_0(\sin^2\theta_c - \sin^2\theta)$. Eqs. (7.9.8) still give the right answers, that is, $\beta_z' = \sqrt{|D_R|}$ and $\alpha_z' = 0$, if $\theta \le \theta_c$, and $\beta_z' = 0$ and $\alpha_z' = \sqrt{|D_R|}$, if $\theta > \theta_c$.

For the case of a very good conductor, we have $\epsilon_I' \gg \epsilon_R'$, or $D_I \gg |D_R|$, and Eqs. (7.9.5) give $\beta_z' \simeq \alpha_z' \simeq \sqrt{D_I/2}$, or

$$\beta_z' \simeq \alpha_z' \simeq \beta' \simeq \alpha' \simeq \sqrt{\frac{\omega\mu_0\sigma}{2}}, \qquad \text{provided} \quad \frac{\sigma}{\omega\epsilon} \gg 1 \tag{7.9.9}$$

In this case, the angle of refraction ϕ for the phase vector $\boldsymbol{\beta}'$ becomes almost zero so that, regardless of the incidence angle θ, the phase planes are almost parallel to the constant-z amplitude planes. Using Eq. (7.9.9), we have:

$$\tan\phi = \frac{k_x}{\beta_z'} = \frac{\omega\sqrt{\mu_0\epsilon}\sin\theta}{\sqrt{\omega\mu_0\sigma/2}} = \sqrt{\frac{2\omega\epsilon}{\sigma}}\sin\theta$$

which is very small regardless of θ. For example, for copper ($\sigma = 5.7\times10^7$ S/m) at 10 GHz, and air on the left side ($\epsilon = \epsilon_0$), we find $\sqrt{2\omega\epsilon/\sigma} = 1.4\times10^{-4}$.

Example 7.9.1: Fig. 7.9.2 shows the TM and TE reflection coefficients as functions of the incident angle θ, for an air-sea water interface at 100 MHz and 1 GHz. For the air side we have $\epsilon = \epsilon_0$ and for the water side: $\epsilon' = 81\epsilon_0 - j\sigma/\omega$, with $\sigma = 4$ S/m, which gives $\epsilon' = (81 - 71.9j)\epsilon_0$ at 1 GHz and $\epsilon' = (81 - 719j)\epsilon_0$ at 100 MHz.

At 1 GHz, we calculate $k' = \omega\sqrt{\mu_0\epsilon'} = \beta' - j\alpha' = 203.90 - 77.45j$ rad/m and $k' = \beta' - j\alpha' = 42.04 - 37.57j$ rad/m at 100 MHz. The following MATLAB code was used to carry out the calculations, using the formulation of this section:

```
ep0 = 8.854e-12; mu0 = 4*pi*1e-7;
sigma = 4; f = 1e9; w = 2*pi*f;
ep1 = ep0; ep2 = 81*ep0 - j*sigma/w;
```

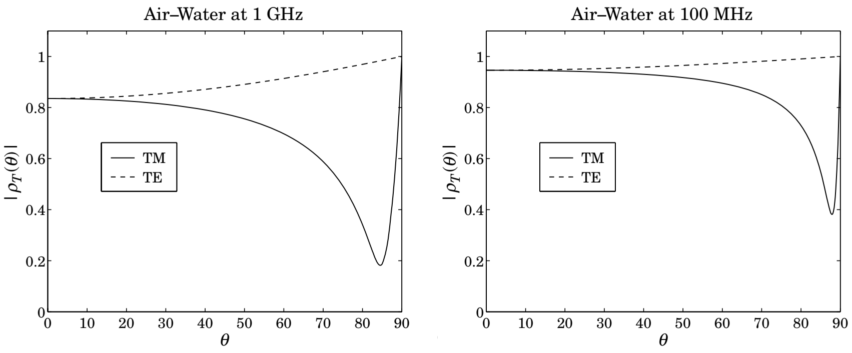


Fig. 7.9.2 TM and TE reflection coefficients for air-water interface.

```
k1 = w*sqrt(mu0*ep1); k2 = w*sqrt(mu0*ep2);              % Eq. (7.9.2)

th = linspace(0,90,901); thr = pi*th/180;

k1x = k1*sin(thr); k1z = k1*cos(thr);
k2z = sqrt(w^2*mu0*ep2 - k1x.^2);                        % Eq. (7.9.6)

rte = abs((k1z - k2z)./(k1z + k2z));                     % Eq. (7.7.2)
rtm = abs((k2z*ep1 - k1z*ep2)./(k2z*ep1 + k1z*ep2));

plot(th,rtm, th,rte);
```

The TM reflection coefficient reaches a minimum at the pseudo-Brewster angles 84.5° and 87.9°, respectively for 1 GHz and 100 MHz.

The reflection coefficients ρ_{TM} and ρ_{TE} can just as well be calculated from Eq. (7.4.2), with $n = 1$ and $n' = \sqrt{\epsilon'/\epsilon_0}$, where for 1 GHz we have $n' = \sqrt{81 - 71.9j} = 9.73 - 3.69j$, and for 100 MHz, $n' = \sqrt{81 - 719j} = 20.06 - 17.92j$. □

In computing the complex square roots in Eq. (7.9.6), MATLAB usually gets the right answer, that is, $\beta'_z \ge 0$ and $\alpha'_z \ge 0$.

If $\epsilon'_R > \epsilon$, then $D_R = \omega^2\mu_0(\epsilon'_R - \epsilon\sin^2\theta)$ is positive for all angles θ, and (7.9.6) may be used without modification for any value of ϵ'_I.

If $\epsilon'_R < \epsilon$ and $\epsilon'_I > 0$, then Eq. (7.9.6) still gives the correct algebraic signs for any angle θ. But when $\epsilon'_I = 0$, that is, for a lossless medium, then $D_I = 0$ and $k'_z = \sqrt{D_R}$. For $\theta > \theta_c$ we have $D_R < 0$ and MATLAB gives $k'_z = \sqrt{D_R} = \sqrt{-|D_R|} = j\sqrt{|D_R|}$, which has the wrong sign for α'_z (we saw that Eqs. (7.9.5) work correctly in this case.)

In order to coax MATLAB to produce the right algebraic sign for α'_z in all cases, we may redefine Eq. (7.9.6) by using double conjugation:

$$k'_z = \beta'_z - j\alpha'_z = \left(\sqrt{(D_R - jD_I)^*}\right)^* = \begin{cases} -j\sqrt{|D_R|}\,, & \text{if } D_I = 0 \text{ and } D_R < 0 \\ \sqrt{D_R - jD_I}\,, & \text{otherwise} \end{cases} \tag{7.9.10}$$

One word of caution, however, is that current versions of MATLAB (ver. ≤ 7.0) may produce inconsistent results for (7.9.10) depending on whether D_I is a scalar or a vector passing through zero. Compare, for example, the outputs from the statements:

```
DI = 0;    kz = conj(sqrt(conj(-1 - j*DI)));
DI = -1:1; kz = conj(sqrt(conj(-1 - j*DI)));
```

Note, however, that Eq. (7.9.10) does work correctly when D_I is a single scalar with D_R being a vector of values, e.g., arising from a vector of angles θ.

Another possible alternative calculation is to add a small negative imaginary part to the argument of the square root, for example with the MATLAB code:

```
kz = sqrt(DR-j*DI-j*realmin);
```

where `realmin` is MATLAB's smallest positive floating point number (typically, equal to 2.2251×10^{-308}). This works well for all cases. Yet, a third alternative is to use Eq. (7.9.6) and then reverse the signs whenever $D_I = 0$ and $D_R < 0$, for example:

```
kz = sqrt(DR-j*DI);
kz(DI==0 & DR<0) = -kz(DI==0 & DR<0);
```

Next, we discuss briefly the energy flux into the lossy medium. It is given by the z-component of the Poynting vector, $\mathcal{P}_z = \frac{1}{2}\hat{\mathbf{z}}\cdot\mathrm{Re}(\boldsymbol{E}\times\boldsymbol{H}^*)$. For the TE case of Eq. (7.7.4), we find at the two sides of the interface:

$$\mathcal{P}_z = \frac{|E_0|^2}{2\omega\mu_0}\,k_z\big(1 - |\rho_{TE}|^2\big)\,,\quad \mathcal{P}'_z = \frac{|E_0|^2}{2\omega\mu_0}\,\beta'_z|\tau_{TE}|^2 e^{-2\alpha'_z z} \tag{7.9.11}$$

where we replaced $\eta_{TE} = \omega\mu_0/k_z$ and $\eta'_{TE} = \omega\mu_0/k'_z$. Thus, the transmitted power attenuates with distance as the wave propagates into the lossy medium.

The two expressions match at the interface, expressing energy conservation, that is, at $z = 0$, we have $\mathcal{P}_z = \mathcal{P}'_z$, which follows from the condition (see Problem 7.7):

$$k_z\big(1 - |\rho_{TE}|^2\big) = \beta'_z|\tau_{TE}|^2 \tag{7.9.12}$$

Because the net energy flow is to the right in the transmitted medium, we must have $\beta'_z \ge 0$. Because also $k_z > 0$, then Eq. (7.9.12) implies that $|\rho_{TE}| \le 1$. For the case of total internal reflection, we have $\beta'_z = 0$, which gives $|\rho_{TE}| = 1$. Similar conclusions can be reached for the TM case of Eq. (7.7.5). The matching condition at the interface is now:

$$\frac{\epsilon}{k_z}\big(1 - |\rho_{TM}|^2\big) = \mathrm{Re}\left(\frac{\epsilon'}{k'_z}\right)|\tau_{TM}|^2 = \frac{\epsilon'_R\beta'_z + \epsilon'_I\alpha'_z}{|k'_z|^2}|\tau_{TM}|^2 \tag{7.9.13}$$

Using the constraint $\omega^2\mu_o\epsilon'_I = 2\beta'_z\alpha'_z$, it follows that the right-hand side will again be proportional to β'_z (with a positive proportionality coefficient.) Thus, the non-negative sign of β'_z implies that $|\rho_{TM}| \le 1$.

7.10 Zenneck Surface Wave

For a lossy medium ϵ', the TM reflection coefficient cannot vanish for any real incident angle θ because the Brewster angle is complex valued: $\tan\theta_B = \sqrt{\epsilon'/\epsilon} = \sqrt{(\epsilon'_R - j\epsilon'_I)/\epsilon}$.

However, ρ_{TM} can vanish if we allow a complex-valued θ, or equivalently, a complex-valued incident wavevector $\boldsymbol{k} = \boldsymbol{\beta} - j\boldsymbol{\alpha}$, even though the left medium is lossless. This leads to the so-called Zenneck surface wave [32,869,870,876,1116].

The corresponding constant phase and amplitude planes in both media are shown in Fig. 7.10.1. On the lossless side, the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are necessarily orthogonal to each other, as discussed in Sec. 2.10.

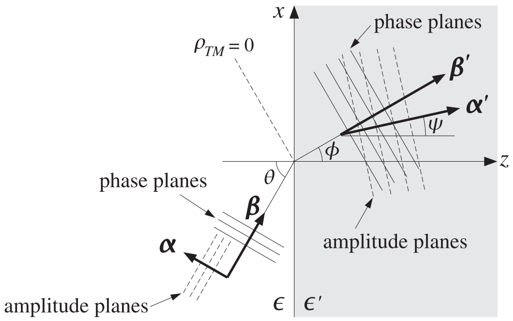


Fig. 7.10.1 Constant-phase and constant-amplitude planes for the Zenneck wave.

We note that the TE reflection coefficient can never vanish (unless $\mu \neq \mu'$) because this would require that $k'_z = k_z$, which together with Snel's law $k'_x = k_x$, would imply that $\boldsymbol{k} = \boldsymbol{k}'$, which is impossible for distinct media.

For the TM case, the fields are given by Eq. (7.7.5) with $\rho_{TM} = 0$ and $\tau_{TM} = 1$. The condition $\rho_{TM} = 0$ requires that $k'_z\epsilon = k_z\epsilon'$, which may be written in the equivalent form $k'_z k^2 = k_z k'^2$. Together with $k_x^2 + k_z^2 = k^2$ and $k_x^2 + k_z'^2 = k'^2$, we have three equations in the three complex unknowns k_x, k_z, k'_z. The solution is easily found to be:

$$k_x = \frac{kk'}{\sqrt{k^2 + k'^2}}\,, \quad k_z = \frac{k^2}{\sqrt{k^2 + k'^2}}\,, \quad k'_z = \frac{k'^2}{\sqrt{k^2 + k'^2}} \tag{7.10.1}$$

where $k = \omega\sqrt{\mu_0\epsilon}$ and $k' = \beta' - j\alpha' = \omega\sqrt{\mu_0\epsilon'}$. These may be written in the form:

$$k_x = \omega\sqrt{\mu_0}\sqrt{\frac{\epsilon\epsilon'}{\epsilon + \epsilon'}}\,, \quad k_z = \omega\sqrt{\mu_0}\,\frac{\epsilon}{\sqrt{\epsilon + \epsilon'}}\,, \quad k'_z = \omega\sqrt{\mu_0}\,\frac{\epsilon'}{\sqrt{\epsilon + \epsilon'}} \tag{7.10.2}$$

Using $k'_x = k_x$, the space-dependence of the fields at the two sides is as follows:

$$e^{-j(k_x x + k_z z)} = e^{-(\alpha_x x + \alpha_z z)}e^{-j(\beta_x x + \beta_z z)}\,, \quad \text{for } z \le 0$$

$$e^{-j(k'_x x + k'_z z)} = e^{-(\alpha_x x + \alpha'_z z)}e^{-j(\beta_x x + \beta'_z z)}\,, \quad \text{for } z \ge 0$$

Thus, in order for the fields not to grow exponentially with distance and to be confined near the

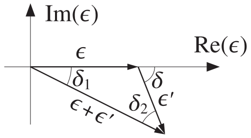
$$\alpha_x > 0\,, \quad \alpha_z < 0\,, \quad \alpha'_z > 0 \tag{7.10.3}$$

These conditions are guaranteed with the sign choices of Eq. (7.10.2). This can be verified by writing

$$\epsilon' = |\epsilon'|e^{-j\delta}$$

$$\epsilon + \epsilon' = |\epsilon + \epsilon'|e^{-j\delta_1}$$

$$\frac{\epsilon'}{\epsilon + \epsilon'} = \left|\frac{\epsilon'}{\epsilon + \epsilon'}\right| e^{-j(\delta - \delta_1)}$$



and noting that $\delta_2 = \delta - \delta_1 > 0$, as follows by inspecting the triangle formed by the three vectors ϵ, ϵ', and $\epsilon + \epsilon'$. Then, the phase angles of k_x, k_z, k'_z are $-\delta_2/2$, $\delta_1/2$, and $-(\delta_2 + \delta_1/2)$, respectively, thus, implying the condition (7.10.3). In drawing this triangle, we made the implicit assumption that $\epsilon'_R > 0$, which is valid for typical lossy dielectrics. In the next section, we discuss surface plasmons for which $\epsilon'_R < 0$.

Although the Zenneck wave attenuates both along the x- and z-directions, the attenuation constant along x tends to be much smaller than that along z. For example, in the weakly lossy approximation, we may write $\epsilon' = \epsilon'_R(1 - j\tau)$, where $\tau = \epsilon'_I/\epsilon'_R \ll 1$ is the loss tangent of ϵ'. Then, we have the following first-order approximations in τ:

$$\sqrt{\epsilon'} = \sqrt{\epsilon'_R}\left(1 - j\frac{\tau}{2}\right), \qquad \frac{1}{\sqrt{\epsilon + \epsilon'}} = \frac{1}{\sqrt{\epsilon + \epsilon'_R}}\left(1 - j\frac{\tau}{2}\frac{\epsilon'_R}{\epsilon + \epsilon'_R}\right)$$

These leads to the first-order approximations for k_x and k_z:

$$k_x = \omega\sqrt{\mu_0}\sqrt{\frac{\epsilon\epsilon'_R}{\epsilon + \epsilon'_R}}\left(1 - j\frac{\tau}{2}\frac{\epsilon}{\epsilon + \epsilon'_R}\right), \qquad k_z = \omega\sqrt{\mu_0}\frac{\epsilon}{\sqrt{\epsilon + \epsilon'_R}}\left(1 + j\frac{\tau}{2}\frac{\epsilon'_R}{\epsilon + \epsilon'_R}\right)$$

It follows that:

$$\alpha_x = \omega\sqrt{\mu_0}\sqrt{\frac{\epsilon\epsilon'_R}{\epsilon + \epsilon'_R}}\,\frac{\tau}{2}\,\frac{\epsilon}{\epsilon + \epsilon'_R}\,, \quad \alpha_z = -\omega\sqrt{\mu_0}\frac{\epsilon}{\sqrt{\epsilon + \epsilon'_R}}\,\frac{\tau}{2}\,\frac{\epsilon'_R}{\epsilon + \epsilon'_R} \quad \Rightarrow \quad \frac{\alpha_x}{|\alpha_z|} = \sqrt{\frac{\epsilon}{\epsilon'_R}}$$

Typically, $\epsilon'_R > \epsilon$, implying that $\alpha_x < |\alpha_z|$. For example, for an air-water interface we have at microwave frequencies $\epsilon'_R/\epsilon = 81$, and for an air-ground interface, $\epsilon'_R/\epsilon = 6$.

If both media are lossless, then both $\boldsymbol{k}$ and $\boldsymbol{k}'$ are real and Eqs. (7.10.1) yield the usual Brewster angle formulas, that is,

$$\tan\theta_B = \frac{k_x}{k_z} = \frac{k'}{k} = \frac{\sqrt{\epsilon'}}{\sqrt{\epsilon}}\,, \quad \tan\theta'_B = \frac{k_x}{k'_z} = \frac{k}{k'} = \frac{\sqrt{\epsilon}}{\sqrt{\epsilon'}}$$

Example 7.10.1: For the data of the air-water interface of Example 7.9.1, we calculate the following Zenneck wavenumbers at 1 GHz and 100 MHz using Eq. (7.10.2):

$f = 1$ GHz	$f = 100$ MHz
$k_x = \beta_x - j\alpha_x = 20.89 - 0.064j$	$k_x = \beta_x - j\alpha_x = 2.1 - 0.001j$
$k_z = \beta_z - j\alpha_z = 1.88 + 0.71j$	$k_z = \beta_z - j\alpha_z = 0.06 + 0.05j$
$k'_z = \beta'_z - j\alpha'_z = 202.97 - 77.80j$	$k'_z = \beta'_z - j\alpha'_z = 42.01 - 37.59j$

The units are in rads/m. As required, α_z is negative. We observe that $\alpha_x \ll |\alpha_z|$ and that the attenuations are much more severe within the lossy medium. □

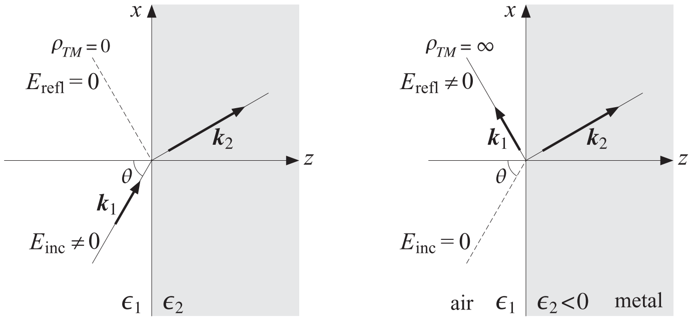
7.11 Surface Plasmons

Consider an interface between two non-magnetic semi-infinite media ϵ_1 and ϵ_2, as shown in Fig. 7.11.1 The wavevectors $\mathbf{k}_1 = \hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_{z1}$ and $\mathbf{k}_2 = \hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_{z2}$ at the two sides must have a common k_x component, as required by Snel's law, and their z-components must satisfy:

$$k_{z1}^2 = k_0^2\varepsilon_1 - k_x^2\,,\quad k_{z2}^2 = k_0^2\varepsilon_2 - k_x^2 \tag{7.11.1}$$

where we defined the *relative* dielectric constants $\varepsilon_1 = \epsilon_1/\epsilon_0$, $\varepsilon_2 = \epsilon_2/\epsilon_0$, and the free-space wavenumber $k_0 = \omega\sqrt{\mu_0\epsilon_0} = \omega/c_0$. The TM reflection coefficient is given by:

$$\rho_{TM} = \frac{k_{z2}\varepsilon_1 - k_{z1}\varepsilon_2}{k_{z2}\varepsilon_1 + k_{z1}\varepsilon_2}$$



Brewster - Zenneck   Surface Plasmons

Fig. 7.11.1 Brewster-Zenneck ($\rho_{TM}=0$) and surface plasmon ($\rho_{TM}=\infty$) cases.

Both the Brewster case for lossless dielectrics and the Zenneck case were characterized by the condition $\rho_{TM} = 0$, or, $k_{z2}\varepsilon_1 = k_{z1}\varepsilon_2$. This condition together with Eqs. (7.11.1) leads to the solution (7.10.2), which is the same in both cases:

$$k_x = k_0\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}\,,\quad k_{z1} = \frac{k_0\varepsilon_1}{\sqrt{\varepsilon_1+\varepsilon_2}}\,,\quad k_{z2} = \frac{k_0\varepsilon_2}{\sqrt{\varepsilon_1+\varepsilon_2}} \tag{7.11.2}$$

Surface plasmons or polaritons are waves that are propagating along the interface and attenuate exponentially perpendicularly to the interface in both media. They are characterized by a *pole* of the reflection coefficient, that is, $\rho_{TM} = \infty$. For such waves to exist, it is necessary to have the conditions:

$$\varepsilon_1\varepsilon_2 < 0 \quad\text{and}\quad \varepsilon_1 + \varepsilon_2 < 0 \tag{7.11.3}$$

at least for the real-parts of these quantities, assuming their imaginary parts are small. If the left medium is an ordinary lossless dielectric $\varepsilon_1 > 0$, such as air, then we must have $\varepsilon_2 < 0$ and more strongly $\varepsilon_2 < -\varepsilon_1$. Conductors, such as silver and gold, have this property for frequencies typically up to ultraviolet. Indeed, using the simple conductivity model (1.9.21), we have for the dielectric constant of a metal:

$$\epsilon(\omega) = \epsilon_0 + \frac{\sigma}{j\omega} = \epsilon_0 + \frac{\epsilon_0\omega_p^2}{j\omega(\omega+\gamma)} \quad\Rightarrow\quad \varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - j\omega\gamma} \tag{7.11.4}$$

Ignoring the imaginary part for the moment, we have

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}$$

which is negative for $\omega < \omega_p$. The plasma frequency is of the order of 1000–2000 THz, and falls in the ultraviolet range. Thus, the condition (7.11.3) is easily met for optical frequencies. If $\varepsilon_1 = 1$, then, the condition $\varepsilon_2 < -\varepsilon_1$ requires further that

$$\varepsilon_2 = 1 - \frac{\omega_p^2}{\omega^2} < -1 \quad\Rightarrow\quad \omega < \frac{\omega_p}{\sqrt{2}}$$

and more generally, $\omega < \omega_p/\sqrt{1+\varepsilon_1}$. The condition $\rho_{TM} = \infty$ means that there is only a "reflected" wave, while the incident field is zero. Indeed, it follows from $E_{\text{refl}} = \rho_{TM}E_{\text{inc}}$, or $E_{\text{inc}} = E_{\text{refl}}/\rho_{TM}$, that E_{inc} will tend to zero for finite E_{refl} and $\rho_{TM}\to\infty$.

The condition $\rho_{TM} = \infty$ is equivalent to the vanishing of the denominator of ρ_{TM}, that is, $k_{z2}\varepsilon_1 = -k_{z1}\varepsilon_2$, which together with Eqs. (7.11.1) leads to a similar solution as (7.10.2), but with a change in sign for k_{z2}:

$$k_x = k_0\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}\,,\quad k_{z1} = \frac{k_0\varepsilon_1}{\sqrt{\varepsilon_1+\varepsilon_2}}\,,\quad k_{z2} = -\frac{k_0\varepsilon_2}{\sqrt{\varepsilon_1+\varepsilon_2}} \tag{7.11.5}$$

The fields at the two sides of the interface are given by Eqs. (7.7.5) by taking the limit $\rho_{TM}\to\infty$ and $\tau_{TM} = 1+\rho_{TM}\to\infty$, which effectively amounts to keeping only the terms that involve ρ_{TM}. The fields have a z-dependence $e^{jk_{z1}z}$ on the left and $e^{-jk_{z2}z}$ on the right, and a common x-dependence e^{-jk_xx}:

$$\left.\begin{aligned}
\mathbf{E}_1 &= E_0\left(\hat{\mathbf{x}} + \frac{k_x}{k_{z1}}\,\hat{\mathbf{z}}\right)e^{jk_{z1}z}\,e^{-jk_xx}\\
\mathbf{H}_1 &= -\hat{\mathbf{y}}\,E_0\,\frac{\omega\epsilon_1}{k_{z1}}\,e^{jk_{z1}z}\,e^{-jk_xx}
\end{aligned}\;\right|\;
\begin{aligned}
\mathbf{E}_2 &= E_0\left(\hat{\mathbf{x}} - \frac{k_x}{k_{z2}}\,\hat{\mathbf{z}}\right)e^{-jk_{z2}z}\,e^{-jk_xx}\\
\mathbf{H}_2 &= \hat{\mathbf{y}}\,E_0\,\frac{\omega\epsilon_2}{k_{z2}}\,e^{-jk_{z2}z}\,e^{-jk_xx}
\end{aligned} \tag{7.11.6}$$

It can be verified easily that these are solutions of Maxwell's equations provided that Eqs. (7.11.1) are satisfied. The boundary conditions are also satisfied. Indeed, the E_x components are the same from both sides, and the conditions $\varepsilon_1E_{z1} = \varepsilon_2E_{z2}$ and $H_{y1} = H_{y2}$ are both equivalent to the pole condition $k_{z2}\varepsilon_1 = -k_{z1}\varepsilon_2$.

The conditions (7.11.3) guarantee that k_x is real and k_{z1}, k_{z_2}, pure imaginary. Setting $\varepsilon_2 = -\varepsilon_{2r}$ with $\varepsilon_{2r} > \varepsilon_1$, we have $\sqrt{\varepsilon_1+\varepsilon_2} = \sqrt{\varepsilon_1-\varepsilon_{2r}} = j\sqrt{\varepsilon_{2r}-\varepsilon_1}$, and $\sqrt{\varepsilon_1\varepsilon_2} = \sqrt{-\varepsilon_1\varepsilon_{2r}} = j\sqrt{\varepsilon_1\varepsilon_{2r}}$. Then, Eqs. (7.11.5) read

$$k_x = k_0\sqrt{\frac{\varepsilon_1\varepsilon_{2r}}{\varepsilon_{2r}-\varepsilon_1}}\,,\quad k_{z1} = -j\,\frac{k_0\varepsilon_1}{\sqrt{\varepsilon_{2r}-\varepsilon_1}}\,,\quad k_{z2} = -j\,\frac{k_0\varepsilon_{2r}}{\sqrt{\varepsilon_{2r}-\varepsilon_1}} \tag{7.11.7}$$

Setting $k_{z1} = -j\alpha_{z1}$ and $k_{z2} = -j\alpha_{z2}$, with both αs positive, the z-dependence at both sides of the interface at $z = 0$ will be:

$$e^{jk_{z1}z} = e^{\alpha_{z1}z} \quad \Big| \quad e^{-jk_{z2}z} = e^{-\alpha_{z2}z}$$

that is, exponentially decaying for both $z < 0$ and $z > 0$. Inserting $\epsilon_{2r} = \omega_p^2/\omega^2 - 1$ into k_x gives the so-called plasmon dispersion relationship, For example, if $\epsilon_1 = 1$,

$$k_x^2 = \frac{\omega^2}{c_0^2}\,\frac{\omega_p^2 - \omega^2}{\omega_p^2 - 2\omega^2}$$

Defining the normalized variables $\bar{\omega} = \omega/\omega_p$ and $\bar{k} = k_x/k_p$, where $k_p = \omega_p/c_0$, we may rewrite the above relationship as,

$$\bar{k}^2 = \bar{\omega}^2\,\frac{1 - \bar{\omega}^2}{1 - 2\bar{\omega}^2}$$

with solution

$$\bar{\omega} = \sqrt{\bar{k}^2 + \frac{1}{2} - \sqrt{\bar{k}^4 + \frac{1}{4}}} \tag{7.11.8}$$

It is depicted in Fig. 7.11.2. In the large k_x limit, it converges to the horizontal line $\omega = \omega_p/\sqrt{2}$. For small k_x, it becomes the dispersion relationship in vacuum, $\omega = c_0 k_x$, which is also depicted in this figure.

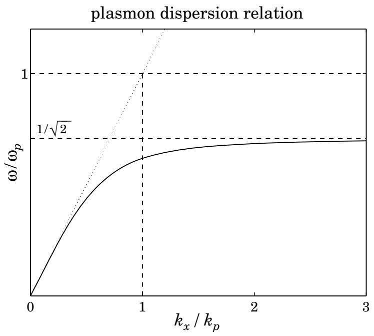


Fig. 7.11.2 Surface plasmon dispersion relationship.

Because the curve stays to the right of the vacuum line $\omega = c_0 k_x$, that is, $k_x > \omega/c_0$, such surface plasmon waves cannot be excited by an impinging plane wave on the interface. However, they can be excited with the help of frustrated total internal reflection, which increases k_x beyond its vacuum value and can match the value of Eq. (7.11.7) resulting into a so-called surface plasmon resonance. We discuss this further in Sec. 8.5.

In fact, the excitation of such plasmon resonance can only take place if the metal side is slightly lossy, that is, when $\epsilon_2 = -\epsilon_{2r} - j\epsilon_{2i}$, with $0 < \epsilon_{2i} \ll \epsilon_{2r}$. In this case, the wavenumber k_x acquires a small imaginary part which causes the gradual attenuation of the wave along the surface, and similarly, k_{z1}, k_{z2}, acquire small real parts. Replacing ϵ_{2r} by $\epsilon_{2r} + j\epsilon_{2i}$ in (7.11.7), we now have:

$$k_x = k_0\sqrt{\frac{\epsilon_1(\epsilon_{2r} + j\epsilon_{2i})}{\epsilon_{2r} + j\epsilon_{2i} - \epsilon_1}}\,,\quad k_{z1} = \frac{-jk_0\epsilon_1}{\sqrt{\epsilon_{2r} + j\epsilon_{2i} - \epsilon_1}}\,,\quad k_{z2} = \frac{-jk_0(\epsilon_{2r} + j\epsilon_{2i})}{\sqrt{\epsilon_{2r} + j\epsilon_{2i} - \epsilon_1}} \tag{7.11.9}$$

Expanding k_x to first-order in ϵ_{2i}, we obtain the approximations:

$$k_x = \beta_x - j\alpha_x\,,\quad \beta_x = k_0\sqrt{\frac{\epsilon_1\epsilon_{2r}}{\epsilon_{2r} - \epsilon_1}}\,,\quad \alpha_x = k_0\left(\frac{\epsilon_1\epsilon_{2r}}{\epsilon_{2r} - \epsilon_1}\right)^{3/2}\frac{\epsilon_{2i}}{2\epsilon_{2r}^2} \tag{7.11.10}$$

Example 7.11.1: Using the value $\epsilon_2 = -16 - 0.5j$ for silver at $\lambda_0 = 632$ nm, and air $\epsilon_1 = 1$, we have $k_0 = 2\pi/\lambda_0 = 9.94$ rad/μm and Eqs. (7.11.9) give the following values for the wavenumbers and the corresponding effective propagation length and penetration depths:

$$k_x = \beta_x - j\alpha_x = 10.27 - 0.0107j \ \text{rad}/\mu\text{m}, \qquad \delta_x = \frac{1}{\alpha_x} = 93.6\ \mu\text{m}$$

$$k_{z1} = \beta_{z1} - j\alpha_{z1} = -0.043 - 2.57j \ \text{rad}/\mu\text{m}, \qquad \delta_{z1} = \frac{1}{\alpha_{z1}} = 390\ \text{nm}$$

$$k_{z2} = \beta_{z2} - j\alpha_{z2} = 0.601 - 41.12j \ \text{rad}/\mu\text{m}, \qquad \delta_{z2} = \frac{1}{\alpha_{z2}} = 24\ \text{nm}$$

Thus, the fields extend more into the dielectric than the metal, but at either side they are confined to distances that are less than their free-space wavelength. □

Surface plasmons, and the emerging field of "plasmonics," are currently active areas of study [553–591] holding promise for the development of nanophotonic devices and circuits that take advantage of the fact that plasmons are confined to smaller spaces than their free-space wavelength and can propagate at decent distances in the nanoscale regime (i.e., tens of μm compared to nm scales.) They are also currently used in chemical and biological sensor technologies, and have other potential medical applications, such as cancer treatments.

7.12 Oblique Reflection from a Moving Boundary

In Sec. 5.8 we discussed reflection and transmission from a moving interface at normal incidence. Here, we present the oblique incidence case. The dielectric medium is assumed to be moving with velocity v perpendicularly to the interface, that is, in the z-direction as shown in Fig. 7.12.1. Other geometries may be found in [435–453].

Let S and S' be the stationary and the moving coordinate frames, whose coordinates $\{t, x, y, z\}$ and $\{t', x', y', z'\}$ are related by the Lorentz transformation of Eq. (H.1) of Appendix H.

We assume a TE plane wave of frequency ω incident obliquely at the moving interface at an angle θ, as measured in the stationary coordinate frame S. Let ω_r, ω_t be the Doppler-shifted frequencies, and θ_r, θ_t, the angles of the reflected and transmitted waves. Because of the motion, these angles no longer satisfy the usual Snel laws of reflection and refraction—however, the do satisfy modified versions of these laws.

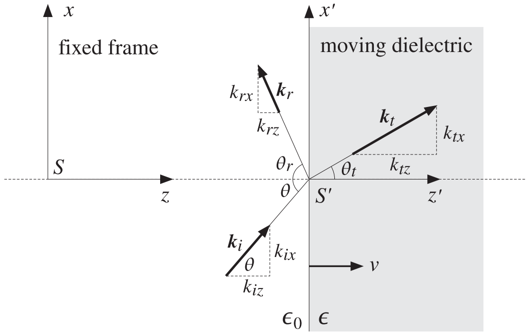


Fig. 7.12.1 Oblique reflection from a moving boundary.

In the moving frame S' with respect to which the dielectric is at rest, we have an ordinary TE oblique incidence problem, solved for example by Eq. (7.7.4), and therefore, all three frequencies will be the same, $\omega' = \omega_r' = \omega_t'$, and the corresponding angles $\theta', \theta_r', \theta_t'$ will satisfy the ordinary Snel laws: $\theta_r' = \theta'$ and $\sin\theta' = n\sin\theta_t'$, where $n = \sqrt{\epsilon/\epsilon_0}$ and the left medium is assumed to be free space.

The electric field has only a y-component and will have the following form at the left and right sides of the interface, in the frame S and in the frame S':

$$\begin{aligned} E_y &= E_i e^{j\phi_i} + E_r e^{j\phi_r}\,, \qquad & E_y &= E_t e^{j\phi_t} \\ E_y' &= E_i' e^{j\phi_i'} + E_r' e^{j\phi_r'}\,, \qquad & E_y' &= E_t' e^{j\phi_t'} \end{aligned} \tag{7.12.1}$$

where $E_r' = \rho_{TE}E_i'$ and $E_t' = \tau_{TE}E_i'$, and from Eq. (7.7.2),

$$\rho_{TE} = \frac{k_{iz}' - k_{tz}'}{k_{iz}' + k_{tz}'} = \frac{\cos\theta' - n\cos\theta_t'}{\cos\theta' + n\cos\theta_t'}\,, \quad \tau_{TE} = 1 + \rho_{TE} = \frac{2\cos\theta'}{\cos\theta' + n\cos\theta_t'} \tag{7.12.2}$$

The propagation phases are Lorentz invariant in the two frames and are given by:

$$\begin{aligned} \phi_i &= \omega t - k_{iz}z - k_{ix}x = \omega' t' - k_{iz}'z' - k_{ix}'x' = \phi_i' \\ \phi_r &= \omega_r t + k_{rz}z - k_{rx}x = \omega' t' + k_{rz}'z' - k_{rx}'x' = \phi_r' \\ \phi_t &= \omega_t t - k_{tz}z - k_{tx}x = \omega' t' - k_{tz}'z' - k_{tx}'x' = \phi_t' \end{aligned} \tag{7.12.3}$$

with incident, reflected, and transmitted wavenumbers given in the frame S' by:

$$\begin{aligned} k_{iz}' &= k_{rz}' = k_i'\cos\theta'\,, \quad & k_{tz}' &= k_t'\cos\theta_t' \\ k_{ix}' &= k_{rx}' = k_i'\sin\theta'\,, \quad & k_{tx}' &= k_t'\sin\theta_t' \end{aligned} \tag{7.12.4}$$

where $k_i' = k_r' = \omega'/c$ and $k_t' = \omega'\sqrt{\epsilon\mu_0} = n\omega'/c$. The relationships between the primed and unprimed frequencies and wavenumbers are obtained by applying the Lorentz transformation (H.14) to the four-vectors $(\omega/c, k_{ix}, 0, k_{iz})$, $(\omega_r, k_{rx}, 0, -k_{rz})$,

and $(\omega_t/c, k_{tx}, 0, k_{tz})$:

$$\begin{aligned} \omega &= \gamma(\omega' + \beta c k_{iz}') = \omega'\gamma(1 + \beta\cos\theta') \\ k_{iz} &= \gamma(k_{iz}' + \frac{\beta}{c}\omega') = \frac{\omega'}{c}\gamma(\cos\theta' + \beta) \\ \omega_r &= \gamma(\omega' - \beta c k_{rz}') = \omega'\gamma(1 - \beta\cos\theta') \\ -k_{rz} &= \gamma(-k_{rz}' + \frac{\beta}{c}\omega') = -\frac{\omega'}{c}\gamma(\cos\theta' - \beta) \\ \omega_t &= \gamma(\omega' + \beta c k_{tz}') = \omega'\gamma(1 + \beta n\cos\theta_t') \\ k_{tz} &= \gamma(k_{tz}' + \frac{\beta}{c}\omega') = \frac{\omega'}{c}\gamma(n\cos\theta_t' + \beta) \end{aligned} \tag{7.12.5}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. Combining Snel's laws for the system S' with the invariance of the x-components of the wavevector under the Lorentz transformation (H.14), we have also:

$$\begin{aligned} & k_{ix} = k_{rx} = k_{tx} = k_{ix}' = k_{rx}' = k_{tx}' \\ & k_i\sin\theta = k_r\sin\theta_r = k_t\sin\theta_t = \frac{\omega'}{c}\sin\theta' = \frac{\omega'}{c}\sin\theta_r' = n\frac{\omega'}{c}\sin\theta_t' \end{aligned} \tag{7.12.6}$$

Because the incident and reflected waves are propagating in free space, their wavenumbers will be $k_i = \omega/c$ and $k_r = \omega_r/c$. This also follows from the invariance of the scalar $(\omega/c)^2 - k^2$ under Lorentz transformations. Indeed, because $k_i' = k_r' = \omega'/c$ in the S' system, we will have:

$$\frac{\omega^2}{c^2} - k_i^2 = \frac{\omega'^2}{c^2} - k_i'^2 = 0\,, \quad \frac{\omega_r^2}{c^2} - k_r^2 = \frac{\omega'^2}{c^2} - k_r'^2 = 0$$

For the transmitted wavenumber k_t, we find from Eqs. (7.12.5) and (7.12.6):

$$k_t = \sqrt{k_{tz}^2 + k_{tx}^2} = \frac{\omega'}{c}\sqrt{\gamma^2(n\cos\theta_t' + \beta)^2 + n^2\sin^2\theta_t'} \tag{7.12.7}$$

Setting $v_t = \omega_t/k_t = c/n_t$, we obtain the "effective" refractive index n_t within the moving dielectric medium:

$$n_t = \frac{c}{v_t} = \frac{ck_t}{\omega_t} = \frac{\sqrt{\gamma^2(n\cos\theta_t' + \beta)^2 + n^2\sin^2\theta_t'}}{\gamma(1 + \beta n\cos\theta_t')} \tag{7.12.8}$$

At normal incidence, this is equivalent to Eq. (5.8.6). Replacing $k_i = \omega/c$, $k_r = \omega_r/c$, and $k_t = \omega_t n_t/c$ in Eq. (7.12.6), we obtain the generalization of Snel's laws:

$$\omega\sin\theta = \omega_r\sin\theta_r = \omega_t n_t\sin\theta_t = \omega'\sin\theta' = \omega'\sin\theta_r' = \omega' n\sin\theta_t' \tag{7.12.9}$$

For a stationary interface, all the frequency factors drop out and we obtain the ordinary Snel laws. The reflected and transmitted frequencies are θ-dependent and are obtained from (7.12.5) by eliminating ω':

$$\omega_r = \omega\frac{1 - \beta\cos\theta'}{1 + \beta\cos\theta'}\,, \quad \omega_t = \omega\frac{1 + \beta n\cos\theta_t'}{1 + \beta\cos\theta'} \tag{7.12.10}$$

Replacing $k_{iz} = k_i\cos\theta = (\omega/c)\cos\theta$ and $k_{rz} = k_r\cos\theta_r = (\omega_r/c)\cos\theta_r$ in Eq. (7.12.5), we obtain the relationship of the angles θ, θ_r to the angle θ':

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'}\,, \quad \cos\theta_r = \frac{\cos\theta' - \beta}{1 - \beta\cos\theta'} \tag{7.12.11}$$

which can also be written as:

$$\cos\theta' = \frac{\cos\theta - \beta}{1 - \beta\cos\theta} = \frac{\cos\theta_r + \beta}{1 + \beta\cos\theta_r} \tag{7.12.12}$$

Solving for θ_r in terms of θ, we obtain:

$$\boxed{\cos\theta_r = \frac{(1+\beta^2)\cos\theta - 2\beta}{1 - 2\beta\cos\theta + \beta^2}} \tag{7.12.13}$$

Inserting $\cos\theta'$ in Eq. (7.12.10), we find the reflected frequency in terms of θ:

$$\boxed{\omega_r = \omega\frac{1 - 2\beta\cos\theta + \beta^2}{1 - \beta^2}} \tag{7.12.14}$$

Eqs. (7.12.13) and (7.12.14) were originally derived by Einstein in his 1905 paper on special relativity [435]. The quantity $n\cos\theta'_t$ can also be written in terms of θ. Using Snel's law and Eq. (7.12.12), we have:

$$n\cos\theta'_t = \sqrt{n^2 - \sin^2\theta'} = \sqrt{n^2 - 1 + \cos^2\theta'} = \sqrt{n^2 - 1 + \left(\frac{\cos\theta - \beta}{1 - \beta\cos\theta}\right)^2}\,, \quad \text{or,}$$

$$n\cos\theta'_t = \frac{\sqrt{(n^2-1)(1-\beta\cos\theta)^2 + (\cos\theta - \beta)^2}}{1 - \beta\cos\theta} \equiv \frac{Q}{1 - \beta\cos\theta} \tag{7.12.15}$$

Using (7.12.15) and the identity $(1+\beta\cos\theta')(1-\beta\cos\theta) = 1 - \beta^2$, we find for the transmitted frequency:

$$\omega_t = \omega\frac{1 + \beta n\cos\theta'_t}{1 + \beta\cos\theta'} = \omega\frac{1 - \beta\cos\theta + \beta Q}{1 - \beta^2} \tag{7.12.16}$$

The TE reflection coefficient (7.12.2) may also be expressed in terms of θ:

$$\rho_{TE} = \frac{\cos\theta' - n\cos\theta'_t}{\cos\theta' + n\cos\theta'_t} = \frac{\cos\theta - \beta - Q}{\cos\theta - \beta + Q} \tag{7.12.17}$$

Next, we determine the reflected and transmitted fields in the frame S. The simplest approach is to apply the Lorentz transformation (H.30) separately to the incident, reflected, and transmitted waves. In the S' frame, a plane wave propagating along the unit vector $\hat{\mathbf{k}}'$ has magnetic field:

$$\boldsymbol{H}' = \frac{1}{\eta}\hat{\mathbf{k}}' \times \boldsymbol{E}' \quad \Rightarrow \quad c\boldsymbol{B}' = c\mu_0\boldsymbol{H}' = \frac{\eta_0}{\eta}\hat{\mathbf{k}}' \times \boldsymbol{E}' = n\hat{\mathbf{k}}' \times \boldsymbol{E}' \tag{7.12.18}$$

where $n = 1$ for the incident and reflected waves. Because we assumed a TE wave and the motion is along the z-direction, the electric field will be perpendicular to the velocity, that is, $\boldsymbol{\beta}\cdot\boldsymbol{E}' = 0$. Using the BAC-CAB rule, Eq. (H.30) then gives:

$$\begin{aligned}\boldsymbol{E} = \boldsymbol{E}_\perp &= \gamma(\boldsymbol{E}'_\perp - \boldsymbol{\beta}\times c\boldsymbol{B}'_\perp) = \gamma(\boldsymbol{E}' - \boldsymbol{\beta}\times c\boldsymbol{B}') = \gamma\big(\boldsymbol{E}' - \boldsymbol{\beta}\times(n\hat{\mathbf{k}}'\times\boldsymbol{E}')\big)\\ &= \gamma\big(\boldsymbol{E}' - n(\boldsymbol{\beta}\cdot\boldsymbol{E}')\hat{\mathbf{k}}' + n(\boldsymbol{\beta}\cdot\hat{\mathbf{k}}')\boldsymbol{E}'\big) = \gamma\boldsymbol{E}'(1 + n\boldsymbol{\beta}\cdot\hat{\mathbf{k}}')\end{aligned} \tag{7.12.19}$$

Applying this result to the incident, reflected, and transmitted fields, we find:

$$\begin{aligned}E_i &= \gamma E'_i(1 + \beta\cos\theta')\\ E_r &= \gamma E'_r(1 - \beta\cos\theta') = \gamma\rho_{TE}E'_i(1 - \beta\cos\theta')\\ E_t &= \gamma E'_t(1 + n\beta\cos\theta'_t) = \gamma\tau_{TE}E'_i(1 + n\beta\cos\theta'_t)\end{aligned} \tag{7.12.20}$$

It follows that the reflection and transmission coefficients will be:

$$\frac{E_r}{E_i} = \rho_{TE}\frac{1 - \beta\cos\theta'}{1 + \beta\cos\theta'} = \rho_{TE}\frac{\omega_r}{\omega}\,, \quad \frac{E_t}{E_i} = \tau_{TE}\frac{1 + n\beta\cos\theta'_t}{1 + \beta\cos\theta'} = \tau_{TE}\frac{\omega_t}{\omega} \tag{7.12.21}$$

The case of a perfect mirror corresponds to $\rho_{TE} = -1$ and $\tau_{TE} = 0$. To be interpretable as a reflection angle, θ_r must be in the range $0 \le \theta_r \le 90^{\mathrm{o}}$, or, $\cos\theta_r > 0$. This requires that the numerator of (7.12.13) be positive, or,

$$(1+\beta^2)\cos\theta - 2\beta \ge 0 \quad \Leftrightarrow \quad \cos\theta \ge \frac{2\beta}{1+\beta^2} \quad \Leftrightarrow \quad \theta \le \operatorname{acos}\Big(\frac{2\beta}{1+\beta^2}\Big) \tag{7.12.22}$$

Because $2\beta/(1+\beta^2) > \beta$, (7.12.22) also implies that $\cos\theta > \beta$, or, $v < c_z = c\cos\theta$. Thus, the z-component of the phase velocity of the incident wave can catch up with the receding interface. At the maximum allowed θ, the angle θ_r reaches 90^{o}. In the above, we assumed that $\beta > 0$. For negative β, there are no restrictions on the range of θ.

7.13 *Geometrical Optics*

Geometrical optics and the concepts of wavefronts and rays can be derived from Maxwell's equations in the short-wavelength or high-frequency limit.

We saw in Chap. 2 that a uniform plane wave propagating in a lossless isotropic dielectric in the direction of a wave vector $\boldsymbol{k} = k\,\hat{\mathbf{k}} = nk_0\,\hat{\mathbf{k}}$ is given by:

$$\boldsymbol{E}(\mathbf{r}) = \boldsymbol{E}_0\,e^{-jnk_0\hat{\mathbf{k}}\cdot\mathbf{r}}\,, \quad \boldsymbol{H}(\mathbf{r}) = \boldsymbol{H}_0\,e^{-jnk_0\hat{\mathbf{k}}\cdot\mathbf{r}}\,, \qquad \hat{\mathbf{k}}\cdot\boldsymbol{E}_0 = 0\,, \quad \boldsymbol{H}_0 = \frac{n}{\eta_0}\hat{\mathbf{k}}\times\boldsymbol{E}_0 \tag{7.13.1}$$

where n is the refractive index of the medium $n = \sqrt{\epsilon/\epsilon_0}$, k_0 and η_0 are the free-space wavenumber and impedance, and $\hat{\mathbf{k}}$, the unit-vector in the direction of propagation.

The wavefronts are defined to be the constant-phase plane surfaces $S(\mathbf{r}) = \text{const.}$, where $S(\mathbf{r}) = n\hat{\mathbf{k}}\cdot\mathbf{r}$. The perpendiculars to the wavefronts are the optical rays.

In an inhomogeneous medium with a space-dependent refractive index $n(\mathbf{r})$, the wavefronts and their perpendicular rays become curved, and can be derived by considering the high-frequency limit of Maxwell's equations. By analogy with Eqs. (7.13.1), we look for solutions of the form:

$$\boldsymbol{E}(\mathbf{r}) = \boldsymbol{E}_0(\mathbf{r})\,e^{-jk_0S(\mathbf{r})}\,, \quad \boldsymbol{H}(\mathbf{r}) = \boldsymbol{H}_0(\mathbf{r})\,e^{-jk_0S(\mathbf{r})} \tag{7.13.2}$$

where we will assume that k_0 is large and that $\boldsymbol{E}_0, \boldsymbol{H}_0$ are slowly-varying functions of $\mathbf{r}$. This means that their space-derivatives are small compared to k_0 or to $1/\lambda$. For example, $|\nabla\times\boldsymbol{E}_0| \ll k_0$.

Inserting these expressions into Maxwell's equations and assuming $\mu = \mu_0$ and $\epsilon = n^2\epsilon_0$, we obtain:

$$\nabla\times\boldsymbol{E} = e^{-jk_0 S}\big(\nabla\times\boldsymbol{E}_0 - jk_0\nabla S\times\boldsymbol{E}_0\big) = -j\omega\mu_0\boldsymbol{H}_0\,e^{-jk_0 S}$$
$$\nabla\times\boldsymbol{H} = e^{-jk_0 S}\big(\nabla\times\boldsymbol{H}_0 - jk_0\nabla S\times\boldsymbol{H}_0\big) = jn^2\omega\epsilon_0\boldsymbol{E}_0\,e^{-jk_0 S}$$

Assuming $|\nabla\times\boldsymbol{E}_0| \ll |k_0\nabla S\times\boldsymbol{E}_0|$, and similarly for $\boldsymbol{H}_0$, and dropping the common phase factor $e^{-jk_0 S}$, we obtain the high-frequency approximations:

$$-jk_0\nabla S\times\boldsymbol{E}_0 = -j\omega\mu_0\boldsymbol{H}_0$$
$$-jk_0\nabla S\times\boldsymbol{H}_0 = jn^2\omega\epsilon_0\boldsymbol{E}_0$$

Replacing $k_0 = \omega\sqrt{\mu_0\epsilon_0}$, and defining the vector $\hat{\mathbf{k}} = \dfrac{1}{n}\nabla S$, we find:

$$\boxed{\boldsymbol{H}_0 = \frac{n}{\eta_0}\,\hat{\mathbf{k}}\times\boldsymbol{E}_0\,,\quad \boldsymbol{E}_0 = -\frac{\eta_0}{n}\,\hat{\mathbf{k}}\times\boldsymbol{H}_0} \tag{7.13.3}$$

These imply the transversality conditions $\hat{\mathbf{k}}\cdot\boldsymbol{E}_0 = \hat{\mathbf{k}}\cdot\boldsymbol{H}_0 = 0$. The consistency of the equations (7.13.3) requires that $\hat{\mathbf{k}}$ be a *unit vector*. Indeed, using the BAC-CAB rule, we have:

$$\hat{\mathbf{k}}\times(\hat{\mathbf{k}}\times\boldsymbol{E}_0) = \hat{\mathbf{k}}(\hat{\mathbf{k}}\cdot\boldsymbol{E}_0) - \boldsymbol{E}_0(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}) = -\boldsymbol{E}_0(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}) = \frac{\eta_0}{n}\hat{\mathbf{k}}\times\boldsymbol{H}_0 = -\boldsymbol{E}_0$$

Thus, we obtain the unit-vector condition, known as the *eikonal equation*:

$$\hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1 \quad\Rightarrow\quad |\nabla S|^2 = n^2 \qquad \text{(eikonal equation)} \tag{7.13.4}$$

This equation determines the wavefront phase function $S(\mathbf{r})$. The rays are the perpendiculars to the constant-phase surfaces $S(\mathbf{r}) = \text{const.}$, so that they are in the direction of ∇S or $\hat{\mathbf{k}}$. Fig. 7.13.1 depicts these wavefronts and rays.

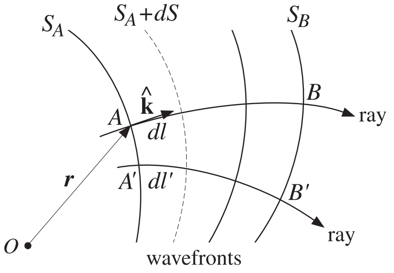


Fig. 7.13.1 Wavefront surfaces and rays.

The ray passing through a point $\mathbf{r}$ on the surface $S(\mathbf{r}) = S_A$, will move ahead by a distance $d\mathbf{r}$ in the direction of the gradient ∇S. The length of $d\mathbf{r}$ is $dl = (d\mathbf{r}\cdot d\mathbf{r})^{1/2}$. The vector $d\mathbf{r}/dl$ is a unit vector in the direction of ∇S and, therefore, it must be equal to $\hat{\mathbf{k}}$. Thus, we obtain the defining equation for the rays:

$$\frac{d\mathbf{r}}{dl} = \hat{\mathbf{k}} \quad\Rightarrow\quad \frac{d\mathbf{r}}{dl} = \frac{1}{n}\nabla S \quad\Rightarrow\quad \boxed{n\,\frac{d\mathbf{r}}{dl} = \nabla S} \tag{7.13.5}$$

The eikonal equation determines S, which in turn determines the rays. The ray equation can be expressed directly in terms of the refractive index by eliminating S. Indeed, differentiating (7.13.5), we have:

$$\frac{d}{dl}\left(n\,\frac{d\mathbf{r}}{dl}\right) = \frac{d}{dl}(\nabla S) = \left(\frac{d\mathbf{r}}{dl}\cdot\nabla\right)\nabla S = \frac{1}{n}(\nabla S\cdot\nabla)\nabla S$$

where, in differentiating along a ray, we used the expression for d/dl:

$$\frac{d}{dl} = \frac{d\mathbf{r}}{dl}\cdot\nabla \tag{7.13.6}$$

But, $\nabla(\nabla S\cdot\nabla S) = 2(\nabla S\cdot\nabla)\nabla S$, which follows from the differential identity Eq. (C.17) of the Appendix. Therefore,

$$\frac{d}{dl}\left(n\,\frac{d\mathbf{r}}{dl}\right) = \frac{1}{n}(\nabla S\cdot\nabla)\nabla S = \frac{1}{2n}\nabla(\nabla S\cdot\nabla S) = \frac{1}{2n}\nabla(n^2) = \frac{1}{2n}2n\nabla n\,,\quad\text{or,}$$

$$\boxed{\frac{d}{dl}\left(n\,\frac{d\mathbf{r}}{dl}\right) = \nabla n} \qquad \text{(ray equation)} \tag{7.13.7}$$

The vectors $\boldsymbol{E}_0, \boldsymbol{H}_0, \hat{\mathbf{k}}$ form a right-handed system as in the uniform plane-wave case. The energy density and flux are:

$$\begin{aligned}
w_e &= \frac{1}{2}\,\mathrm{Re}\big[\tfrac{1}{2}\epsilon\,\boldsymbol{E}\cdot\boldsymbol{E}^*\big] = \frac{1}{4}\epsilon_0 n^2|\boldsymbol{E}_0|^2\\
w_m &= \frac{1}{4}\mu_0|\boldsymbol{H}_0|^2 = \frac{1}{4}\mu_0\,\frac{n^2}{\eta_0^2}\,|\boldsymbol{E}_0|^2 = \frac{1}{4}\epsilon_0 n^2|\boldsymbol{E}_0|^2 = w_e\\
w &= w_e + w_m = \frac{1}{2}\epsilon_0 n^2|\boldsymbol{E}_0|^2\\
\boldsymbol{\mathcal{P}} &= \frac{1}{2}\,\mathrm{Re}\big[\boldsymbol{E}\times\boldsymbol{H}^*\big] = \frac{n}{2\eta_0}\,\hat{\mathbf{k}}\,|\boldsymbol{E}_0|^2
\end{aligned} \tag{7.13.8}$$

Thus, the energy transport velocity is:

$$\boldsymbol{v} = \frac{\boldsymbol{\mathcal{P}}}{w} = \frac{c_0}{n}\,\hat{\mathbf{k}} \tag{7.13.9}$$

The velocity $\boldsymbol{v}$ depends on $\mathbf{r}$, because n and $\hat{\mathbf{k}}$ do.

7.14 Fermat's Principle

An infinitesimal movement by dl along a ray will change the wavefront phase function by $dS = ndl$. Indeed, using Eq. (7.13.6) and the eikonal equation we find:

$$\frac{dS}{dl} = \frac{d\mathbf{r}}{dl}\cdot\nabla S = \frac{1}{n}\nabla S\cdot\nabla S = \frac{1}{n}n^2 = n \tag{7.14.1}$$

Integrating *along a ray path* from a point A on wavefront $S(\mathbf{r}) = S_A$ to a point B on wavefront $S(\mathbf{r}) = S_B$, as shown in Fig. 7.13.1, gives rise to the net phase change:

$$\boxed{S_B - S_A = \int_A^B dS = \int_A^B ndl} \tag{7.14.2}$$

The right-hand side is recognized as the *optical path length* from A to B. It is proportional to the travel time of moving from A to B with the ray velocity $\boldsymbol{v}$ given by Eq. (7.13.9). Indeed, we have $dl = \boldsymbol{v}\cdot\hat{\mathbf{k}}\,dt = c_0\,dt/n$, or, $dS = ndl = c_0dt$. Thus,

$$\boxed{S_B - S_A = \int_A^B ndl = c_0\int_{t_A}^{t_B} dt = c_0(t_B - t_A)} \tag{7.14.3}$$

Fermat's Principle states that among all possible paths connecting the two points A and B, the geometrical optics ray path is the one the *minimizes* the optical path length (7.14.3), or equivalently, the travel time between the two points. The solution to this minimization problem is the ray equation (7.13.7).

Any path connecting the points A and B may be specified parametrically by the curve $\mathbf{r}(\tau)$, where the parameter τ varies over an interval $\tau_A \le \tau \le \tau_B$. The length dl may be written as:

$$dl = (d\mathbf{r}\cdot d\mathbf{r})^{1/2} = (\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{1/2}d\tau\,, \quad \text{where} \quad \dot{\mathbf{r}} = \frac{d\mathbf{r}}{d\tau} \tag{7.14.4}$$

Then, the functional to be minimized is:

$$\int_A^B ndl = \int_{\tau_A}^{\tau_B}\mathcal{L}(\mathbf{r},\dot{\mathbf{r}})\,d\tau\,, \quad \text{where} \quad \mathcal{L}(\mathbf{r},\dot{\mathbf{r}}) = n(\mathbf{r})\,(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{1/2} \tag{7.14.5}$$

The minimization of Eq. (7.14.5) may be viewed as a problem in variational calculus with Lagrangian function $\mathcal{L}$. Its solution is obtained from the Euler-Lagrange equations:

$$\frac{d}{d\tau}\left(\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{r}}}\right) = \frac{\partial\mathcal{L}}{\partial\mathbf{r}} \tag{7.14.6}$$

See [827–829] for a review of such methods. The required partial derivatives are:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{r}} = \frac{\partial n}{\partial\mathbf{r}}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{1/2}\,, \qquad \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{r}}} = n\,\dot{\mathbf{r}}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{-1/2} = n\frac{d\mathbf{r}}{d\tau}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{-1/2}$$

The Euler-Lagrange equations are then:

$$\frac{d}{d\tau}\left(n\frac{d\mathbf{r}}{d\tau}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{-1/2}\right) = \frac{\partial n}{\partial\mathbf{r}}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{1/2} \quad \text{or,}$$

$$\boxed{(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{-1/2}\frac{d}{d\tau}\left(n\frac{d\mathbf{r}}{d\tau}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{-1/2}\right) = \frac{\partial n}{\partial\mathbf{r}}} \tag{7.14.7}$$

Using $dl = (\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})^{1/2}d\tau$, we may rewrite these in terms of the length variable dl, resulting in the same equations as (7.13.7), that is,

$$\frac{d}{dl}\left(n\frac{d\mathbf{r}}{dl}\right) = \frac{\partial n}{\partial\mathbf{r}} \tag{7.14.8}$$

A variation of Fermat's principle states that the phase change between two wavefront surfaces is *independent* of the choice of the ray path taken between the surfaces. Following a different ray between points A' and B', as shown in Fig. 7.13.1, gives the same value for the net phase change as between the points A and B:

$$\boxed{S_B - S_A = \int_A^B ndl = \int_{A'}^{B'} ndl'} \tag{7.14.9}$$

This form is useful for deriving the shapes of parabolic reflector and hyperbolic lens antennas discussed in Chap. 18.

It can also be used to derive Snel's law of reflection and refraction. Fig. 7.14.1 shows the three families of incident, reflected, and refracted plane wavefronts on a horizontal interface between media n_a and n_b, such that the incident, reflected, and refracted rays are perpendicular to their corresponding wavefronts.

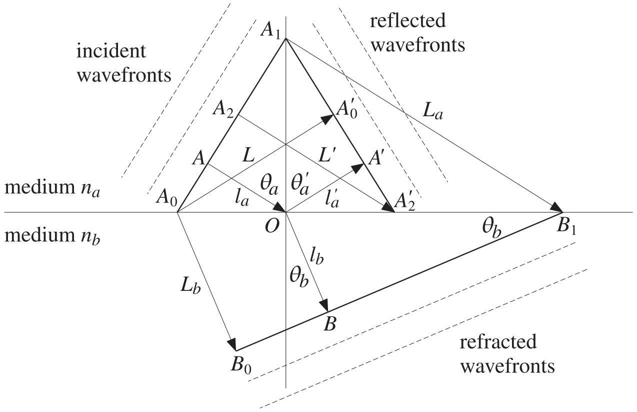


Fig. 7.14.1 Snel's laws of reflection and refraction.

For the reflection problem, we consider the ray paths between the wavefront surfaces A_0A_1 and A_1A_2'. Fermat's principle implies that the optical path length of the rays AOA', A_0A_0', and A_2A_2' will be the same. This gives the condition:

$$n_a(l_a + l_a') = n_aL = n_aL' \quad\Rightarrow\quad L = L'$$

where L and L' are the lengths of the rays A_0A_0 and A_2A_2'. It follows that the two triangles $A_0A_2A_2'$ and $A_0A_0'A_2'$ will be congruent. and therefore, their angles at the vertices A_0 and A_2' will be equal. Thus, $\theta_a = \theta_a'$.

For the refraction problem, we consider the ray paths AOB, A_0B_0, and A_1B_1 between the wavefronts A_0A_1 and B_0B_1. The equality of the optical lengths gives now:

$$n_a l_a + n_b l_b = n_b L_b = n_a L_a \quad\Rightarrow\quad \frac{L_a}{L_b} = \frac{n_b}{n_a}$$

But, the triangles $A_0A_1B_1$ and $A_0B_0B_1$ have a common base A_0B_1. Therefore,

$$\frac{L_a}{L_b} = \frac{\sin\theta_a}{\sin\theta_b}$$

Thus, we obtain Snel's law of refraction:

$$\frac{L_a}{L_b} = \frac{\sin\theta_a}{\sin\theta_b} = \frac{n_b}{n_a} \quad\Rightarrow\quad n_a\sin\theta_a = n_b\sin\theta_b$$

7.15 Ray Tracing

In this section, we apply Fermat's principle of least optical path to derive the ray curves in several integrable examples of inhomogeneous media.

As a special case of Eq. (7.14.8), we consider a stratified half-space $z \ge 0$, shown in Fig. 7.15.1, in which the refractive index is a function of z, but not of x.

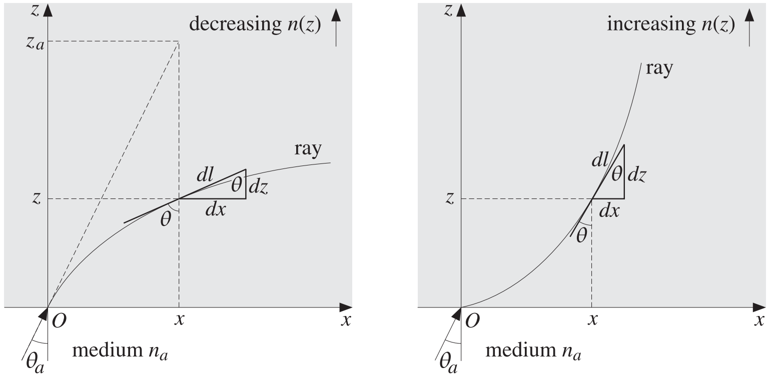


Fig. 7.15.1 Rays in an inhomogeneous medium.

Let θ be the angle formed by the tangent on the ray at point (x,z) and the vertical. Then, we have from the figure $dx = dl\sin\theta$ and $dz = dl\cos\theta$. Because $\partial n/\partial x = 0$, the ray equation (7.14.8) applied to the x-coordinate reads:

$$\frac{d}{dl}\left(n\frac{dx}{dl}\right) = 0 \quad\Rightarrow\quad n\frac{dx}{dl} = \text{const.} \quad\Rightarrow\quad n\sin\theta = \text{const.} \tag{7.15.1}$$

This is the generalization of Snel's law to an inhomogeneous medium. The constant may be determined by evaluating it at the entry point $z = 0$ and $x = 0$. We take the constant to be $n_a\sin\theta_a$. Thus, we write (7.15.2) as:

$$\boxed{n(z)\sin\theta(z) = n_a\sin\theta_a} \qquad \text{(generalized Snel's law)} \tag{7.15.2}$$

The z-component of the ray equation is, using $dz = dl\cos\theta$:

$$\frac{d}{dl}\left(n\frac{dz}{dl}\right) = \frac{dn}{dz} \quad\Rightarrow\quad \cos\theta\,\frac{d}{dz}\,(n\cos\theta) = \frac{dn}{dz} \tag{7.15.3}$$

This is a consequence of Eq. (7.15.2). To see this, we write:

$$n\cos\theta = \sqrt{n^2 - n^2\sin^2\theta} = \sqrt{n^2 - n_a^2\sin^2\theta_a} \tag{7.15.4}$$

Differentiating it with respect to z, we obtain Eq. (7.15.3). The ray in the left Fig. 7.15.1 is bending away from the z-axis with an increasing angle $\theta(z)$. This requires that $n(z)$ be a *decreasing* function of z. Conversely, if $n(z)$ is *increasing* as in the right figure, then $\theta(z)$ will be decreasing and the ray will curve towards the z-axis.

Thus, we obtain the rule that a ray always bends in the direction of increasing $n(z)$ and away from the direction of decreasing $n(z)$.

The constants n_a and θ_a may be taken to be the launch values at the origin, that is, $n(0)$ and $\theta(0)$. Alternatively, if there is a discontinuous change between the lower and upper half-spaces, we may take n_a, θ_a to be the refractive index and incident angle from below.

The ray curves can be determined by relating x and z. From Fig. 7.15.1, we have $dx = dz\tan\theta$, which in conjunction with Eqs. (7.15.2) and (7.15.4) gives:

$$\frac{dx}{dz} = \tan\theta = \frac{n\sin\theta}{n\cos\theta} = \frac{n_a\sin\theta_a}{\sqrt{n^2(z) - n_a^2\sin^2\theta_a}} \tag{7.15.5}$$

Integrating, we obtain:

$$\boxed{x = \int_0^z \frac{n_a\sin\theta_a}{\sqrt{n^2(z') - n_a^2\sin^2\theta_a}}\,dz'} \qquad \text{(ray curve)} \tag{7.15.6}$$

An object at the point (x,z) will appear to an observer sitting at the entry point O as though it is at the apparent location (x, z_a), as shown in Fig. 7.15.1. The apparent or virtual height will be $z_a = x\cot\theta_a$, which can be combined with Eq. (7.15.6) to give:

$$\boxed{z_a = \int_0^z \frac{n_a\cos\theta_a}{\sqrt{n^2(z') - n_a^2\sin^2\theta_a}}\,dz'} \qquad \text{(virtual height)} \tag{7.15.7}$$

The length z_a can be greater or less than z. For example, if the upper half-space is homogeneous with $n_b < n_a$, then $z_a > z$. If $n_b > n_a$, then $z_a < z$, as was the case in Example 7.5.4.

Next, we discuss a number of examples in which the integral (7.15.6) can be done explicitly to derive the ray curves.

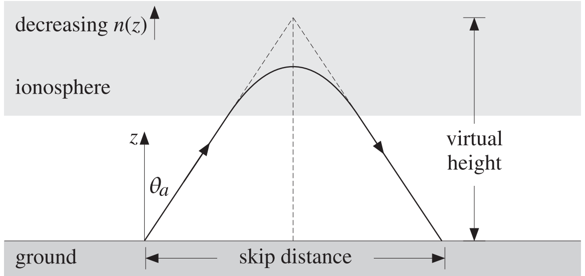


Fig. 7.15.2 Ionospheric refraction.

Example 7.15.1: *Ionospheric Refraction.* Radio waves of frequencies typically in the range of about 4–40 MHz can be propagated at large distances such as 2000–4000 km by bouncing off the ionosphere. Fig. 7.15.2 depicts the case of a flat ground.

The atmosphere has a typical extent of 600 km and is divided in layers: the *troposphere* up to 10 km, the *stratosphere* at 10–50 km, and the *ionosphere* at 50–600 km. The ionosphere is further divided in sublayers, such as the D, E, F_1, and F_2 layers at 50–100 km, 100–150 km, 150–250 km, and 250–400 km, respectively.

The ionosphere consists mostly of ionized nitrogen and oxygen at low pressure. The ionization is due to solar radiation and therefore it varies between night and day. We recall from Sec. 1.9 that a collisionless plasma has an effective refractive index:

$$n^2 = \frac{\epsilon(\omega)}{\epsilon_0} = 1 - \frac{\omega_p^2}{\omega^2}\,, \qquad \omega_p^2 = \frac{Ne^2}{\epsilon_0 m} \tag{7.15.8}$$

The electron density N varies with the time of day and with height. Typically, N increases through the D and E layers and reaches a maximum value in the F layer, and then decreases after that because, even though the solar radiation is more intense, there are fewer gas atoms to be ionized.

Thus, the ionosphere acts as a stratified medium in which $n(z)$ first decreases with height from its vacuum value of unity and then it increases back up to unity. We will indicate the dependence on height by rewriting Eq. (7.15.8) in the form:

$$n^2(z) = 1 - \frac{f_p^2(z)}{f^2}\,, \qquad f_p^2(z) = \frac{N(z)e^2}{4\pi^2\epsilon_0 m} \tag{7.15.9}$$

If the wave is launched straight up and its frequency f is larger than the largest f_p, then it will penetrate through the ionosphere and be lost. But, if there is a height such that $f = f_p(z)$, then at that height $n(z) = 0$ and the wave will be reflected back down.

If the wave is launched at an angle θ_a, then it follows from Snel's law that while the refractive index $n(z)$ is decreasing, the angle of refraction $\theta(z)$ will be increasing and the ray path will bend more and more away from z-axis as shown on the left of Fig. 7.15.1. Below the ionosphere, we may assume that the atmosphere has refractive index $n_a = 1$. Then, the angle $\theta(z)$ may be written as:

$$\sin^2\theta(z) = \frac{n_a^2\sin^2\theta_a}{n^2(z)} = \frac{\sin^2\theta_a}{1 - \dfrac{f_p^2(z)}{f^2}} \tag{7.15.10}$$

Because $\sin\theta(z)$ is required to be less than unity, we obtain the restriction:

$$\sin^2\theta_a \le 1 - \frac{f_p^2(z)}{f^2} \quad\Rightarrow\quad f_p(z) \le f\cos\theta_a \tag{7.15.11}$$

If there is a height, say $z_{\max}$, at which this becomes an equality, $f_p(z_{\max}) = f\cos\theta_a$, then Eq. (7.15.10) would imply that $\sin\theta(z_{\max}) = 1$, or that $\theta(z_{\max}) = 90^{\text{o}}$. At that height, the ray is horizontal and it will proceed to bend downwards, effectively getting reflected from the ionosphere.

If f is so large that Eq. (7.15.11) is satisfied only as a strict inequality, then the wave will escape through all the layers of the ionosphere. Thus, there is a maximum frequency, the so called *maximum usable frequency* (MUF), that will guarantee a reflection. There is also a *lowest usable frequency* (LUF) below which there is too much absorption of the wave, such as in the D layer, to be reflected at sufficient strength for reception.

As an oversimplified, but analytically tractable, model of the ionosphere we assume that the electron density increases linearly with height, up to a maximal height $z_{\max}$. Thus, the quantities $f_p^2(z)$ and $n^2(z)$ will also depend linearly on height:

$$f_p^2(z) = f_{\max}^2\frac{z}{z_{\max}}\,, \qquad n^2(z) = 1 - \frac{f_{\max}^2}{f^2}\frac{z}{z_{\max}}\,, \quad \text{for} \quad 0 \le z \le z_{\max} \tag{7.15.12}$$

Over the assumed height range $0 \le z \le z_{\max}$, the condition (7.15.11) must also be satisfied. This restricts further the range of z. We have:

$$f_p^2(z) = f_{\max}^2\frac{z}{z_{\max}} \le f^2\cos^2\theta_a \quad\Rightarrow\quad \frac{z}{z_{\max}} \le \frac{f^2\cos^2\theta_a}{f_{\max}^2} \tag{7.15.13}$$

If the right-hand side is greater than unity, so that $f\cos\theta_a > f_{\max}$, then there is no height z at which (7.15.11) achieves an equality, and the wave will escape. But, if $f\cos\theta_a \le f_{\max}$, then there is height, say z_0, at which the ray bends horizontally, that is,

$$\frac{z_0}{z_{\max}} = \frac{f^2\cos^2\theta_a}{f_{\max}^2} \quad\Rightarrow\quad z_0 = \frac{z_{\max}f^2\cos^2\theta_a}{f_{\max}^2} \tag{7.15.14}$$

The condition $f\cos\theta_a \le f_{\max}$ can be written as $f \le f_{\text{MUF}}$, where the MUF is in this case, $f_{\text{MUF}} = f_{\max}/\cos\theta_a$. The integral (7.15.6) can be done explicitly resulting in:

$$x = \frac{2z_{\max}\sin^2\theta_a}{a^2}\left[\cos\theta_a - \sqrt{\cos^2\theta_a - a^2\frac{z}{z_{\max}}}\,\right] \tag{7.15.15}$$

where we defined $a = f_{\max}/f$. Solving for z in terms of x, we obtain:

$$z - z_0 = -\frac{1}{4F}(x - x_0)^2 \tag{7.15.16}$$

where

$$x_0 = \frac{2z_{\max}\sin\theta_a\cos\theta_a}{a^2}\,, \quad F = \frac{z_{\max}\sin^2\theta_a}{a^2}$$

Therefore, the ray follows a downward parabolic path with vertex at (x_0, z_0) and focal length F, as shown in Fig. 7.15.3. □

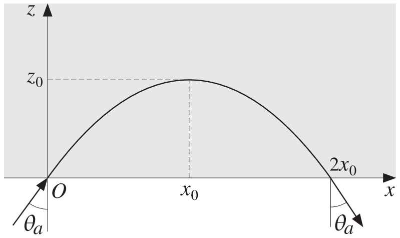


Fig. 7.15.3 Parabolic ray.

Example 7.15.2: *Mirages.* Temperature gradients can cause several types of mirage effects that are similar to ionospheric refraction. On a hot day, the ground is warmer than the air above it and therefore, the refractive index of the air is lower at the ground than a short distance above. (Normally, the air pressure causes the refractive index to be highest at the ground, decreasing with height.)

Because $n(z)$ decreases downwards, a horizontal ray from an object near the ground will initially be refracted downwards, but then it will bend upwards again and may arrive at an observer as though it were coming from below the ground, causing a mirage. Fig. 7.15.4 depicts a typical case. The ray path is like the ionospheric case, but inverted.

Such mirages are seen in the desert and on highways, which appear wet at far distances. Various types of mirages are discussed in [49–51,1303].

As a simple integrable model, we may assume that $n(z)$ increases linearly with height z, that is, $n(z)= n_0 + \kappa z$, where κ is the rate of increase per meter. For heights near the ground, this implies that $n^2(z)$ will also increase linearly:

$$\boxed{n(z)= n_0 + \kappa z} \quad \Rightarrow \quad \boxed{n^2(z)= n_0^2 + 2n_0\kappa z} \tag{7.15.17}$$

We consider a ray launched at a downward angle θ_a from an object with (x, z) coordinates $(0, h)$, as shown. Let $n_a^2 = n_0^2 + 2n_0\kappa h$ be the refractive index at the launch height. For convenience, we assume that the observer is also at height h. Because the ray will travel downward to points $z < h$, and then bend upwards, we integrate the ray equation over the limits $[z, h]$ and find:

$$x = \int_z^h \frac{n_a \sin\theta_a}{\sqrt{n^2(z') - n_a^2 \sin^2\theta_a}}\, dz' = \frac{n_a \sin\theta_a}{n_0\kappa}\left[n_a\cos\theta_a - \sqrt{n_a^2\cos^2\theta_a + 2n_0\kappa(z-h)}\right]$$

where we used the approximation $n^2(z)= n_0^2 + 2n_0\kappa z$ in the integral. Solving for z in terms of x, we obtain the parabolic ray:

$$z = h + \frac{x(x-2x_0)}{4F}\,, \qquad x_0 = \frac{d}{2} = \frac{n_a^2\sin\theta_a\cos\theta_a}{n_0\kappa}\,, \quad F = \frac{n_a^2\sin^2\theta_a}{2n_0\kappa}$$

where d is the distance to the observer and F is the focal length. The apex of the parabola is at $x = x_0 = d/2$ at a height z_0 given by:

$$z_0 = h - \frac{x_0^2}{4F} \quad \Rightarrow \quad z - z_0 = \frac{1}{4F}(x-x_0)^2$$

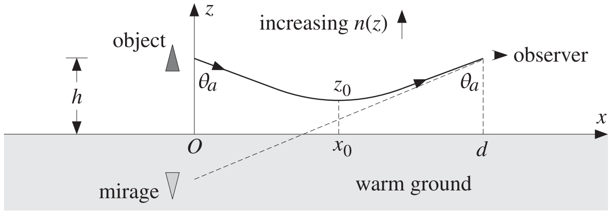


Fig. 7.15.4 Mirage due to a temperature gradient.

The launch angle that results in the ray being tangential to ground is obtained by setting the apex height to zero, $z_0 = 0$. This gives a condition that may be solved for θ_a:

$$x_0 = \sqrt{4Fh} \quad \Rightarrow \quad \sin\theta_a = \frac{n_0}{n_a} \quad \Rightarrow \quad F = \frac{n_0}{2\kappa} \quad \Rightarrow \quad \boxed{x_0 = \sqrt{\frac{2hn_0}{\kappa}}} \tag{7.15.18}$$

The corresponding $d = 2x_0$ is the maximum distance of the observer from the object for which a ray can just touch the ground. □

Example 7.15.3: *Atmospheric Refraction* [49–51]. Because of the compression of gravity, the density of the atmosphere[†] and its refractive index n are highest near the ground and decrease exponentially with height. A simplified model [680], which assumes a uniform temperature and constant acceleration of gravity, is as follows:

$$n(z)= 1 + (n_0 - 1)e^{-z/h_c} \tag{7.15.19}$$

The refractive index on the ground is approximately $n_0 = 1.0003$ (it also has some dependence on wavelength, which we ignore here.) The characteristic height h_c is given by $h_c = RT/Mg$, where R, T, M, g are the universal gas constant, temperature in absolute units, molecular mass of the atmosphere and acceleration of gravity:

$$R = 8.31\,\frac{\text{J}}{\text{K mole}}\,, \quad M = 0.029\,\frac{\text{kg}}{\text{mole}}\,, \quad g = 9.8\,\frac{\text{m}}{\text{s}^2}$$

For a temperature of $T = 303$K, or 30 °C, we find a height of $h_c = 8.86$ km. At a height of a few h_c, the refractive index becomes unity.

The bending of the light rays as they pass through the atmosphere cause the apparent displacement of a distant object, such as a star, the sun, or a geosynchronous satellite. Fig. 7.15.5 illustrates this effect. The object appears to be closer to the zenith.

[†] The troposphere and some of the stratosphere, consisting mostly of molecular nitrogen and oxygen.

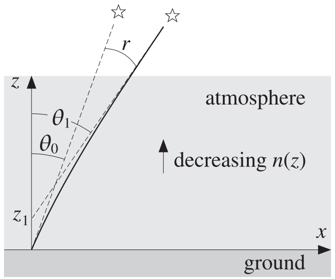


Fig. 7.15.5 Atmospheric refraction.

The look-angle θ_0 at the ground and the true angle of the object θ_1 are related by Snel's law $n_1\sin\theta_1 = n_0\sin\theta_0$. But at large distances (many multiples of h_c), we have $n_1 = 1$. Therefore,

$$\sin\theta_1 = n_0\sin\theta_0 \tag{7.15.20}$$

The refraction angle is $r = \theta_1 - \theta_0$. Assuming a small r, we may use the approximation $\sin(\theta_0 + r) = \sin\theta_0 + r\cos\theta_0$. Then, Eq. (7.15.20) gives the approximate expression:

$$r = (n_0 - 1)\tan\theta_0$$

The maximum viewing angle in this model is such that $n_0\sin\theta_0 = \sin\theta_1 = 1$, corresponding to $\theta_1 = 90^o$ and $\theta_0 = \text{asin}(1/n_0) = 88.6^o$, for $n_0 = 1.0003$.

The model assumes a flat Earth. When the curvature of the Earth is taken into account, the total atmospheric refraction near the horizon, that is, near $\theta_0 = 90^o$, is about 0.65^o for a sea-level observer [49]. The setting sun subtends an angle of about 0.5^o. Therefore, when it appears about to set and its lower edge is touching the horizon, it has already moved *below* the horizon.

The model of Eq. (7.15.19) may be integrated exactly. The ray curves are obtained from Eq. (7.15.6). Setting $n_a = n_0$, $\theta_a = \theta_0$ and using the definition (7.15.20), we obtain:

$$x = h_c\tan\theta_1\left[\text{atanh}\left(\frac{A}{B}\right) - \text{atanh}\left(\frac{A_0}{B_0}\right)\right] = \tan\theta_1\left[z + h_c\,\ln\left(\frac{A+B}{A_0+B_0}\right)\right] \tag{7.15.21}$$

where the quantities A, B, A_0, B_0 are defined as follows:

$$A = n(z) - \sin^2\theta_1\,, \qquad A_0 = n_0 - \sin^2\theta_1$$
$$B = \cos\theta_1\sqrt{n^2(z) - \sin^2\theta_1}\,, \qquad B_0 = \cos\theta_1\sqrt{n_0^2 - \sin^2\theta_1}$$

Thus, A_0, B_0 are the values of A, B at $z = 0$. It can be shown that $A > B$ and therefore, the hyperbolic arc-tangents will be complex-valued. However, the difference of the two atanh terms is real and can be transformed into the second expression in (7.15.21) with the help of the result $A^2 - B^2 = (A_0^2 - B_0^2)e^{-2z/h_c}$.

In the limit of $z \gg h_c$, the quantities A, B tend to $A_1 = B_1 = \cos^2\theta_1$. and the ray equation becomes the *straight line* with a slope of $\tan\theta_1$:

$$x = (z + z_1)\tan\theta_1\,, \quad z_1 = h_c\,\ln\left(\frac{A_1+B_1}{A_0+B_0}\right) \tag{7.15.22}$$

This asymptotic line is depicted in Fig. 7.15.5, intercepting the z-axis at an angle of θ_1. □

Example 7.15.4: *Bouguer's Law.* The previous example assumed a flat Earth. For a spherical Earth in which the refractive index is a function of the radial distance r only, that is, $n(r)$, the ray tracing procedure must be modified.

Snel's law $n(z)\sin\theta(z) = n_0\sin\theta_0$ must be replaced by *Bouguer's law* [598], which states that the quantity $rn(r)\sin\theta$ remain constant:

$$\boxed{rn(r)\sin\theta(r) = r_0 n(r_0)\sin\theta_0} \qquad \text{(Bouguer's law)} \tag{7.15.23}$$

where $\theta(r)$ is the angle of the tangent to the ray and the radial vector. This law can be derived formally by considering the ray equations in spherical coordinates and assuming that $n(r)$ depends only on r [828].

A simpler derivation is to divide the atmosphere in equal-width spherical layers and assume that the refractive index is homogeneous in each layer. In Fig. 7.15.6, the layers are defined by the radial distances and refractive indices r_i, n_i, $i = 0, 1, 2, \dots$.

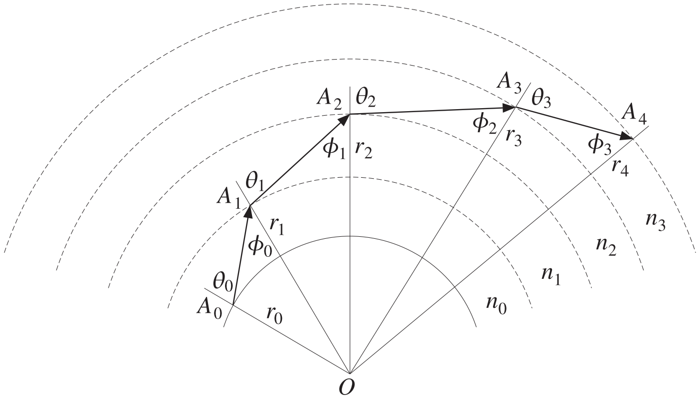


Fig. 7.15.6 Ray tracing in spherically stratified medium.

For sufficiently small layer widths, the ray segments between the points $A_0, A_1, A_2, \dots$ are tangential to the radial circles. At the interface point A_3, Snel's law gives $n_2\sin\phi_2 = n_3\sin\theta_3$. On the other hand, from the triangle OA_2A_3, we have the law of sines:

$$\frac{r_2}{\sin\phi_2} = \frac{r_3}{\sin(\pi - \theta_2)} = \frac{r_3}{\sin\theta_2} \quad\Rightarrow\quad r_2\sin\theta_2 = r_3\sin\phi_2$$

Combining with Snel's law, we obtain:

$$r_2 n_2\sin\theta_2 = r_3 n_2\sin\phi_2 = r_3 n_3\sin\theta_3$$

Thus, the product $r_i n_i\sin\theta_i$ is the same for all $i = 0, 1, 2, \dots$. Defining an effective refractive index by $n_{\text{eff}}(r) = n(r)r/r_0$, Bouguer's law may be written as Snel's law:

$$n_{\text{eff}}(r)\sin\theta(r) = n_0\sin\theta_0$$

where we have the initial value $n_{\text{eff}}(r_0) = n_0 r_0/r_0 = n_0$. □

Example 7.15.5: *Standard Atmosphere over Flat Earth.* For radiowave propagation over ground, the International Telecommunication Union (ITU) [836,837] defines a “standard” atmosphere with the values $n_0 = 1.000315$ and $h_c = 7.35$ km, in Eq. (7.15.19).

For heights of about one kilometer, such that $z \ll h_c$, we may linearize the exponential, $e^{-z/h_c} = 1 - z/h_c$, and obtain the refractive index for the standard atmosphere:

$$\boxed{n(z) = n_0 - \kappa z}\,, \qquad \kappa = \frac{n_0 - 1}{h_c} = \frac{315\times 10^{-6}}{7.35\times 10^{3}} = 4.2857\times 10^{-8}\ \text{m}^{-1} \tag{7.15.24}$$

This is similar to Eq. (7.15.17), with the replacement $\kappa \to -\kappa$. Therefore, we expect the rays to be parabolic bending downwards as in the case of the ionosphere. A typical ray between two antennas at height h and distance d is shown in Fig. 7.15.7.

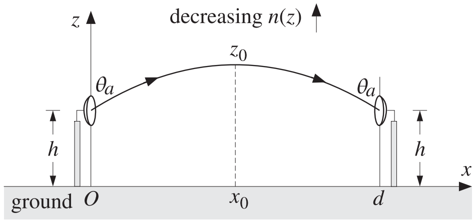


Fig. 7.15.7 Rays in standard atmosphere over a flat Earth.

Assuming an upward launch angle θ_a and defining the refractive index n_a at height h through $n_a^2 = n_0^2 - 2n_0\kappa h$, we obtain the ray equations by integrating over $[h, z]$:

$$x = \int_h^z \frac{n_a \sin\theta_a}{\sqrt{n^2(z') - n_a^2\sin^2\theta_a}}\,dz' = \frac{n_a\sin\theta_a}{n_0\kappa}\Big[n_a\cos\theta_a - \sqrt{n_a^2\cos^2\theta_a - 2n_0\kappa(z-h)}\Big]$$

where we used $n^2(z) = n_0^2 - 2n_0\kappa z$. Solving for z, we obtain the parabola:

$$z = h - \frac{x(x - 2x_0)}{4F}\,, \qquad x_0 = \frac{d}{2} = \frac{n_a^2\sin\theta_a\cos\theta_a}{n_0\kappa}\,, \quad F = \frac{n_a^2\sin^2\theta_a}{2n_0\kappa}$$

where d is the distance to the observer and F is the focal length. The apex of the parabola is at $x = x_0 = d/2$ at a height z_0 given by:

$$z_0 = h + \frac{x_0^2}{4F} \quad\Rightarrow\quad z - z_0 = -\frac{1}{4F}\,(x - x_0)^2$$

The minus sign in the right-hand side corresponds to a downward parabola with apex at the point (x_0, z_0). □

Example 7.15.6: *Standard Atmosphere over Spherical Earth.* We saw in Example 7.15.4 that in Bouguer’s law the refractive index $n(r)$ may be replaced by an effective index $n_e(r) = n(r)r/r_0$. Applying this to the case of the Earth with $r_0 = R$ and $r = R + z$, where R is the Earth radius and z the height above the surface, we have $n_e(z) = n(z)(R+z)/R$, or,

$$n_e(z) = n(z)\Big(1 + \frac{z}{R}\Big) = (n_0 - \kappa z)\Big(1 + \frac{z}{R}\Big)$$

Thus, the spherical Earth introduces the factor $(1 + z/R)$, which increases with height and counteracts the decreasing $n(z)$. Keeping only linear terms in z, we find:

$$\boxed{n_e(z) = n_0 + \kappa_e z}\,, \qquad \kappa_e = \frac{n_0}{R} - \kappa \tag{7.15.25}$$

For the average Earth radius $R = 6370$ km and the ITU values of n_0 and κ given in Eq. (7.15.24), we find that the effective κ_e is *positive*:

$$\kappa_e = 1.1418\times 10^{-7}\ \text{m}^{-1} \tag{7.15.26}$$

Making the approximation $n^2(z) = n_0^2 + 2n_0\kappa_e z$ will result in parabolic rays bending upwards as in Example 7.15.2.

Often, an equivalent Earth radius is defined by $\kappa_e = n_0/R_e$ so that the effective refractive index may be assumed to arise only from the curvature of the equivalent Earth:

$$n_e(z) = n_0 + \kappa_e z = n_0\Big(1 + \frac{z}{R_e}\Big)$$

In units of R, we have:

$$\boxed{\frac{R_e}{R} = \frac{n_0}{\kappa_e R} = \frac{n_0}{n_0 - \kappa R} = 1.3673} \tag{7.15.27}$$

which is usually replaced by $R_e = 4R/3$. In this model, the refractive index is assumed to be uniform above the surface of the equivalent Earth, $n(z) = n_0$.

The ray paths are determined by considering only the geometrical effect of the spherical surface. For example, to determine the maximum distance x_0 at which a ray from a transmitter at height h just grazes the ground, we may either use the results of Eq. (7.15.18), or consider a straight path that is tangential to the equivalent Earth, as shown in Fig. 7.15.8.

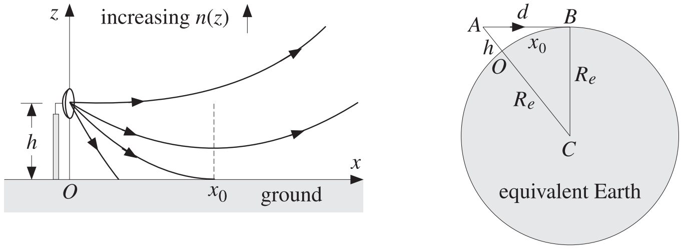


Fig. 7.15.8 Rays over a spherical Earth.

Setting $\kappa_e = n_0/R_e$ in Eq. (7.15.18), we obtain:

$$\boxed{x_0 = \sqrt{\frac{2n_0 h}{\kappa_e}} = \sqrt{2hR_e}} \tag{7.15.28}$$

On the other hand, because $h \ll R_e$ the arc length $x_0 = (OB)$ may be taken to be a straight line in Fig. 7.15.8. Applying the Pythagorean theorem to the two orthogonal triangles OAB and CAB we find that:

$$x_0^2 + h^2 = d^2 = (h + R_e)^2 - R_e^2 = h^2 + 2hR_e \quad \Rightarrow \quad x_0^2 = 2hR_e$$

which is the same as Eq. (7.15.28). □

Example 7.15.7: *Graded-Index Optical Fibers.* In Example 7.5.5, we considered a step-index optical fiber in which the rays propagate by undergoing total internal reflection bouncing off the cladding walls. Here, we consider a graded-index fiber in which the refractive index of the core varies radially from the center value n_f to the cladding value n_c at the edge of the core. Fig. 7.15.9 shows the geometry.

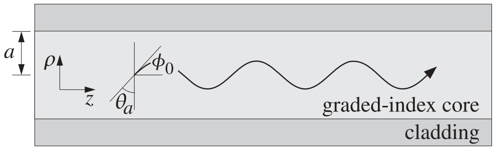


Fig. 7.15.9 Graded-index optical fiber.

As a simple model, we assume a parabolic dependence on the radial distance. We may write in cylindrical coordinates, where a is the radius of the core:

$$n^2(\rho) = n_f^2\left(1 - \Delta^2\,\frac{\rho^2}{a^2}\right), \qquad \Delta^2 = \frac{n_f^2 - n_c^2}{n_f^2} \tag{7.15.29}$$

Inserting this expression into Eq. (7.15.6), and changing variables from z, x to ρ, z, the integral can be done explicitly resulting in:

$$z = \frac{a\sin\theta_a}{\Delta}\,\operatorname{asin}\left(\frac{\rho\Delta}{a\cos\theta_a}\right) \tag{7.15.30}$$

Inverting the arc-sine, we may solve for ρ in terms of z obtaining the following sinusoidal variation of the radial coordinate, where we also changed from the incident angle θ_a to the initial launch angle $\phi_0 = 90^\circ - \theta_a$:

$$\boxed{\rho = \frac{\tan\phi_0}{\kappa}\,\sin(\kappa z)}, \qquad \kappa = \frac{\Delta}{a\cos\phi_0} \tag{7.15.31}$$

For small launch angles ϕ_0, the oscillation frequency becomes independent of ϕ_0, that is, $\kappa = \Delta/(a\cos\phi_0) \simeq \Delta/a$. The rays described by Eq. (7.15.31) are meridional rays, that is, they lie on a plane through the fiber axis, such as the xz- or yz-plane.

There exist more general ray paths that have nontrivial azimuthal dependence and propagate in a helical fashion down the guide [830–835]. □

7.16 Snel's Law in Negative-Index Media

Consider the planar interface between a normal (i.e., positive-index) lossless medium ϵ, μ and a lossless negative-index medium [353] ϵ', μ' with negative permittivity and permeability, $\epsilon' < 0$ and $\mu' < 0$, and negative refractive index $n' = -\sqrt{\mu'\epsilon'/\mu_0\epsilon_0}$. The refractive index of the left medium is as usual $n = \sqrt{\mu\epsilon/\mu_0\epsilon_0}$. A TE or TM plane wave is incident on the interface at an angle θ, as shown in Fig. 7.16.1.

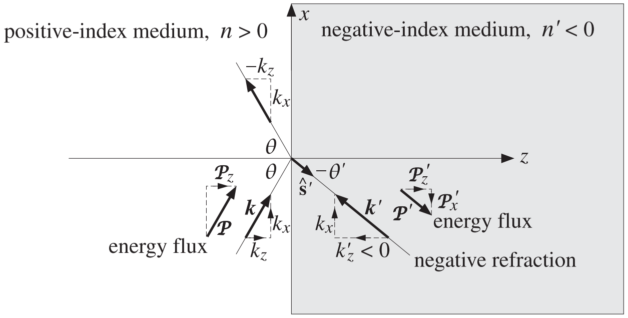


Fig. 7.16.1 Refraction into a negative-index medium.

Because $n' < 0$, Snel's law implies that the refracted ray will bend in the opposite direction (e.g., with a negative refraction angle) than in the normal refraction case. This follows from:

$$n\sin\theta = n'\sin\theta' = -|n'|\sin\theta' = |n'|\sin(-\theta') \tag{7.16.1}$$

As a result, the wave vector $\boldsymbol{k}'$ of the refracted wave will point *towards* the interface, instead of away from it. Its x-component matches that of the incident wave vector $\boldsymbol{k}$, that is, $k'_x = k_x$, which is equivalent to Snel's law (7.16.1), while its z-component points towards the interface or the negative z-direction in the above figure.

Formally, we have $\boldsymbol{k}' = k'\hat{\mathbf{s}}'$, where $\hat{\mathbf{s}}'$ is the unit vector in the direction of the refracted ray pointing away from the interface, and $k' = -\omega\sqrt{\mu'\epsilon'} = n'k_0$, with k_0 the free-space wavenumber $k_0 = \omega\sqrt{\mu_0\epsilon_0} = \omega/c_0$. As we see below, the energy flux Poynting vector $\boldsymbol{\mathcal{P}}'$ of the refracted wave is opposite $\boldsymbol{k}'$ and points in the direction of $\hat{\mathbf{s}}'$, and therefore, carries energy away from the interface. Thus, component-wise we have:

$$k'_x = n'k_0\sin\theta' = k_x = nk_0\sin\theta\,, \quad k'_z = n'k_0\cos\theta' = -|n'|k_0\cos\theta' < 0$$

The TE and TM wave solutions at both sides of the interface are still given by Eqs. (7.7.4) and (7.7.5), and reproduced below (with $e^{j\omega t}$ suppressed):

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= \hat{\mathbf{y}}\,E_0\big[e^{-jk_z z} + \rho_{TE}\,e^{jk_z z}\big]e^{-jk_x x} \\
\boldsymbol{H}(\mathbf{r}) &= \frac{E_0}{\eta_{TE}}\left[\left(-\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z z} + \rho_{TE}\left(\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z z}\right]e^{-jk_x x} \\
\boldsymbol{E}'(\mathbf{r}) &= \hat{\mathbf{y}}\,\tau_{TE}\,E_0 e^{-jk'_z z}e^{-jk_x x} \\
\boldsymbol{H}'(\mathbf{r}) &= \frac{\tau_{TE}\,E_0}{\eta'_{TE}}\left(-\hat{\mathbf{x}} + \frac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{-jk'_z z}e^{-jk_x x}
\end{aligned}} \qquad \text{(TE)} \qquad (7.16.2)$$

where, allowing for magnetic media, we have

$$\eta_{TE} = \frac{\omega\mu}{k_z}\,,\quad \eta'_{TE} = \frac{\omega\mu'}{k'_z}\,,\quad \rho_{TE} = \frac{\eta'_{TE} - \eta_{TE}}{\eta'_{TE} + \eta_{TE}} = \frac{k_z\mu' - k'_z\mu}{k_z\mu' + k'_z\mu}\,,\quad \tau_{TE} = 1 + \rho_{TE} \qquad (7.16.3)$$

For the TM case we have:

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= E_0\left[\left(\hat{\mathbf{x}} - \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z z} + \rho_{TM}\left(\hat{\mathbf{x}} + \frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z z}\right]e^{-jk_x x} \\
\boldsymbol{H}(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{E_0}{\eta_{TM}}\big[e^{-jk_z z} - \rho_{TM}\,e^{jk_z z}\big]e^{-jk_x x} \\
\boldsymbol{E}'(\mathbf{r}) &= \tau_{TM}\,E_0\left(\hat{\mathbf{x}} - \frac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{-jk'_z z}e^{-jk_x x} \\
\boldsymbol{H}'(\mathbf{r}) &= \hat{\mathbf{y}}\,\frac{\tau_{TM}\,E_0}{\eta'_{TM}}e^{-jk'_z z}e^{-jk_x x}
\end{aligned}} \qquad \text{(TM)} \qquad (7.16.4)$$

with

$$\eta_{TM} = \frac{k_z}{\omega\epsilon}\,,\quad \eta'_{TM} = \frac{k'_z}{\omega\epsilon'}\,,\quad \rho_{TM} = \frac{\eta'_{TM} - \eta_{TM}}{\eta'_{TM} + \eta_{TM}} = \frac{k'_z\epsilon - k_z\epsilon'}{k'_z\epsilon + k_z\epsilon'}\,,\quad \tau_{TM} = 1 + \rho_{TM} \qquad (7.16.5)$$

One can verify easily that in both cases the above expressions satisfy Maxwell's equations and the boundary conditions at the interface, provided that

$$\begin{aligned}
k_x^2 + k_z^2 &= \omega^2\mu\epsilon = n^2 k_0^2 \\
k_x^2 + k_z'^2 &= \omega^2\mu'\epsilon' = n'^2 k_0^2
\end{aligned} \qquad (7.16.6)$$

In fact, Eqs. (7.16.2)–(7.16.6) describe the most general case of arbitrary, homogeneous, isotropic, positive- or negative-index, and possibly lossy, media on the left and right and for either propagating or evanescent waves. We concentrate, next, on the case when the left medium is a positive-index lossless medium, $\mu > 0$ and $\epsilon > 0$, and the right one is lossless with $\mu' < 0$ and $\epsilon' < 0$, and consider a propagating incident wave with $k_x = nk_0\sin\theta$ and $k_z = nk_0\cos\theta$ and assume, for now, that $n \le |n'|$ to avoid evanescent waves into the right medium. The Poynting vector $\boldsymbol{\mathcal{P}}'$ in the right medium can be calculated from Eqs. (7.16.2) and (7.16.4):

$$\begin{aligned}
\text{(TE):}\quad \boldsymbol{\mathcal{P}}' &= \frac{1}{2}\,\mathrm{Re}\big(\boldsymbol{E}' \times \boldsymbol{H}'^*\big) = \frac{1}{2}|\tau_{TE}|^2|E_0|^2\left[\hat{\mathbf{z}}\,\mathrm{Re}\left(\frac{k'_z}{\omega\mu'}\right) + \hat{\mathbf{x}}\,\mathrm{Re}\left(\frac{k_x}{\omega\mu'}\right)\right] \\
\text{(TM):}\quad \boldsymbol{\mathcal{P}}' &= \frac{1}{2}\,\mathrm{Re}\big(\boldsymbol{E}' \times \boldsymbol{H}'^*\big) = \frac{1}{2}|\tau_{TM}|^2|E_0|^2\left[\hat{\mathbf{z}}\,\mathrm{Re}\left(\frac{\omega\epsilon'}{k'_z}\right) + \hat{\mathbf{x}}\,\mathrm{Re}\left(\frac{\omega\epsilon' k_x}{|k'_z|^2}\right)\right]
\end{aligned} \qquad (7.16.7)$$

Because $\mu' < 0$ and $\epsilon' < 0$, and k'_z is real, the requirement of positive energy flux away from the interface, $\mathcal{P}'_z > 0$, requires that $k'_z < 0$ in both cases. Similarly, because $k_x > 0$, the x-component of $\boldsymbol{\mathcal{P}}'$ will be negative, $\mathcal{P}'_x < 0$. Thus, the vector $\boldsymbol{\mathcal{P}}'$ has the direction shown in Fig. 7.16.1. We note also that the z-component is preserved across the interface, $\mathcal{P}_z = \mathcal{P}'_z$. This follows from the relationships:

$$\begin{aligned}
\mathcal{P}_z &= \frac{1}{2}|E_0|^2\frac{k_z}{\omega\mu}\big(1 - |\rho_{TE}|^2\big) = \frac{1}{2}|E_0|^2|\tau_{TE}|^2\,\mathrm{Re}\left(\frac{k'_z}{\omega\mu'}\right) = \mathcal{P}'_z \\
\mathcal{P}_z &= \frac{1}{2}|E_0|^2\frac{\omega\epsilon}{k_z}\big(1 - |\rho_{TM}|^2\big) = \frac{1}{2}|E_0|^2|\tau_{TM}|^2\,\mathrm{Re}\left(\frac{\omega\epsilon'}{k'_z}\right) = \mathcal{P}'_z
\end{aligned} \qquad (7.16.8)$$

If $n > |n'|$, the possibility of total internal reflection arises. When $\sin\theta > |n'|/n$, then $k_z'^2 = n'^2k_0^2 - k_x^2 = k_0^2(n'^2 - n^2\sin^2\theta)$ is negative and k'_z becomes pure imaginary. In this case, the real-parts in the right-hand side of Eq. (7.16.8) are zero, showing that $|\rho_{TE}| = |\rho_{TM}| = 1$ and there is no (time-averaged) power flow into the right medium.

For magnetic media, including negative-index media, the Brewster angle may also exist for TE polarization, corresponding to $\rho_{TE} = 0$. This condition is equivalent to $k'_z\mu = k_z\mu'$. Similarly $\rho_{TM} = 0$ is equivalent to $k'_z\epsilon = k_z\epsilon'$. These two conditions imply the following relationship for the Brewster angles:

$$\begin{aligned}
\rho_{TE} = 0 \quad &\Rightarrow \quad k'_z\mu = k_z\mu' \quad \Rightarrow \quad (\mu'^2 - \mu^2)\sin^2\theta_B = \mu'^2 - \frac{\mu'\epsilon'}{\mu\epsilon}\,\mu^2 \\
\rho_{TM} = 0 \quad &\Rightarrow \quad k'_z\epsilon = k_z\epsilon' \quad \Rightarrow \quad (\epsilon'^2 - \epsilon^2)\sin^2\theta_B = \epsilon'^2 - \frac{\mu'\epsilon'}{\mu\epsilon}\,\epsilon^2
\end{aligned} \qquad (7.16.9)$$

Clearly, these may or may not have a solution, such that $0 < \sin^2\theta_B < 1$, depending on the relative values of the constitutive parameters. For non-magnetic media, $\mu = \mu' = \mu_0$, the TE case has no solution and the TM case reduces to the usual expression:

$$\sin^2\theta_B = \frac{\epsilon'^2 - \epsilon'\epsilon}{\epsilon'^2 - \epsilon^2} = \frac{\epsilon'}{\epsilon' + \epsilon} = \frac{n'^2}{n'^2 + n^2}$$

Assuming that ϵ, μ and ϵ', μ' have the same sign (positive or negative), we may replace these quantities with their absolute values in Eq. (7.16.9). Defining the parameters $x = |\mu'/\mu|$ and $y = |\epsilon'/\epsilon|$, we may rewrite (7.16.9) in the form:

$$\begin{aligned}
\text{TE case:}\quad & \left(1 - \frac{1}{x^2}\right)\sin^2\theta_B = \left(1 - \frac{y}{x}\right) \\
\text{TM case:}\quad & \left(1 - \frac{1}{y^2}\right)\sin^2\theta_B = \left(1 - \frac{x}{y}\right)
\end{aligned} \qquad (7.16.10)$$

with the TE and TM cases being obtained from each other by the duality transformations $x \to y$ and $y \to x$. It is straightforward to verify that the ranges of the x, y parameters for which a Brewster angle exists are as follows:

$$\begin{aligned}
\text{TE case:}\quad & x > 1\,,\ y < x\,,\ y > \frac{1}{x}\,,\quad \text{or,}\quad x < 1\,,\ y > x\,,\ y < \frac{1}{x} \\
\text{TM case:}\quad & y > 1\,,\ x < y\,,\ y > \frac{1}{x}\,,\quad \text{or,}\quad y < 1\,,\ x > y\,,\ y < \frac{1}{x}
\end{aligned} \qquad (7.16.11)$$

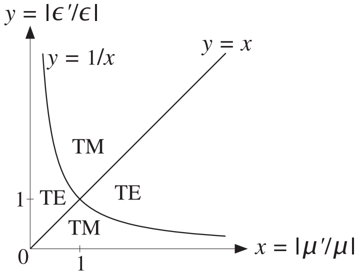


Fig. 7.16.2 Brewster angle regions.

These regions [657], which are bounded by the curves $y = x$ and $y = 1/x$, are shown in Fig. 7.16.2. We note, in particular, that the TE and TM regions are non-overlapping.

The unusual property of Snel's law in negative-index media that the refracted ray bends in the opposite direction than in the normal case has been verified experimentally in artificial metamaterials constructed by arrays of wires and split-ring resonators [359], and by transmission line elements [392–394,414,427]. Another consequence of Snel's law is the possibility of a perfect lens [360] in the case $n' = -1$. We discuss this in Sec. 8.6.

7.17 Problems

7.1 The matching of the tangential components of the electric and magnetic fields resulted in Snel's laws and the matching matrix Eq. (7.3.11). In both the TE and TM polarization cases, show that the remaining boundary conditions $B_z = B_z'$ and $D_z = D_z'$ are also satisfied.

7.2 Show that the Fresnel coefficients (7.4.2) may be expressed in the forms:

$$\rho_{TM} = \frac{\sin 2\theta' - \sin 2\theta}{\sin 2\theta' + \sin 2\theta} = \frac{\tan(\theta' - \theta)}{\tan(\theta' + \theta)}\,, \quad \rho_{TE} = \frac{\sin(\theta' - \theta)}{\sin(\theta' + \theta)}$$

7.3 Show that the refractive index ratio n'/n can be expressed in terms of the ratio $r = \rho_{TM}/\rho_{TE}$ and the incident angle θ by:

$$\frac{n'}{n} = \sin\theta\left[1 + \left(\frac{1+r}{1-r}\right)^2 \tan^2\theta\right]^{1/2}$$

This provides a convenient way of measuring the refractive index n' from measurements of the Fresnel coefficients [676]. It is valid also for complex n'.

7.4 It is desired to design a Fresnel rhomb such that the exiting ray will be elliptically polarized with relative phase difference ϕ between its TE and TM components. Let $\sin\theta_c = 1/n$ be the critical angle within the rhomb. Show that the rhomb angle replacing the 54.6° angle in Fig. 7.5.6 can be obtained from:

$$\sin^2\theta = \frac{\cos^2\theta_c \pm \sqrt{\cos^4\theta_c - 4\sin^2\theta_c\tan^2(\phi/4)}}{2\tan^2(\phi/4) + \cos^2\theta_c \pm \sqrt{\cos^4\theta_c - 4\sin^2\theta_c\tan^2(\phi/4)}}$$

Show ϕ is required to satisfy $\tan(\phi/4) \le (n - n^{-1})/2$.

7.5 Show the relationship (7.9.7) for the ratio ρ_{TM}/ρ_{TE} by first proving and then using the following identities in the notation of Eq. (7.7.4):

$$(k_z' \pm k_z)(k_x^2 \pm k_z k_z') = k^2 k_z' \pm k'^2 k_z$$

Using (7.9.7), show that when both media are lossless, the ratio ρ_{TM}/ρ_{TE} can be expressed directly in terms of the angles of incidence and refraction, θ and θ':

$$\frac{\rho_{TM}}{\rho_{TE}} = \frac{\cos(\theta + \theta')}{\cos(\theta - \theta')}$$

Using this result argue that $|\rho_{TM}| \le |\rho_{TE}|$ at all angles θ. Argue also that $\theta_B + \theta_B' = 90^\circ$, for the Brewster angles. Finally, show that for lossless media with $\epsilon > \epsilon'$, and angles of incidence $\theta \ge \theta_c$, where $\sin\theta_c = \sqrt{\epsilon'/\epsilon}$, we have:

$$\frac{\rho_{TM}}{\rho_{TE}} = \frac{j\sqrt{\sin^2\theta - \sin^2\theta_c} + \sin\theta\tan\theta}{j\sqrt{\sin^2\theta - \sin^2\theta_c} - \sin\theta\tan\theta}$$

Explain how this leads to the design equation (7.5.8) of the Fresnel rhomb.

7.6 Let the incident, reflected, and transmitted waves at an interface be:

$$\mathbf{E}_+(\mathbf{r}) = \mathbf{E}_+ e^{-j\mathbf{k}_+\cdot\mathbf{r}}\,, \quad \mathbf{E}_-(\mathbf{r}) = \mathbf{E}_- e^{-j\mathbf{k}_-\cdot\mathbf{r}}\,, \quad \mathbf{E}'(\mathbf{r}) = \mathbf{E}_0' e^{-j\mathbf{k}'\cdot\mathbf{r}}$$

where $\mathbf{k}_\pm = k_x\hat{\mathbf{x}} \pm k_z\hat{\mathbf{z}}$ and $\mathbf{k}' = k_x\hat{\mathbf{x}} + k_z'\hat{\mathbf{z}}$. Show that the reflection and transmission coefficients defined in Eqs. (7.7.1)–(7.7.5) can be summarized compactly by the following vectorial relationships, which are valid for both the TE and TM cases:

$$\frac{\mathbf{k}_\pm \times (\mathbf{E}_0' \times \mathbf{k}_\pm)}{k^2} = \frac{2k_z}{k_z \pm k_z'}\,\mathbf{E}_\pm$$

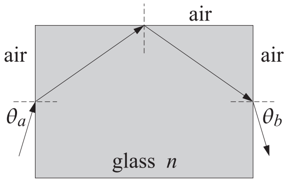
7.7 Using Eqs. (7.7.4), derive the expressions (7.9.11) for the Poynting vectors. Derive similar expressions for the TM case.

Using the definitions in Eqs. (7.3.12), show that if the left medium is lossless and the right one lossy, the following relationship holds:

$$\frac{1}{\eta_T}\left(1 - |\rho_T|^2\right) = \mathrm{Re}\left(\frac{1}{\eta_T'}\right)|\tau_T|^2$$

Then, show that Eqs. (7.9.12) and (7.9.13) are special cases of this result, specialized to the TE and TM cases.

7.8 A light ray enters a glass block from one side, suffers a total internal reflection from the top side, and exits from the opposite side, as shown below. The glass refractive index is $n = 1.5$.



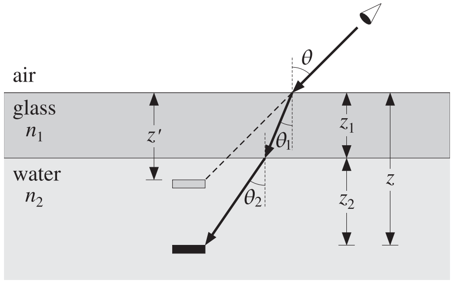
a. How is the exit angle θ_b related to the entry angle θ_a? Explain.

b. Show that all rays, regardless of the entry angle θ_a, will suffer total internal reflection at the top side.

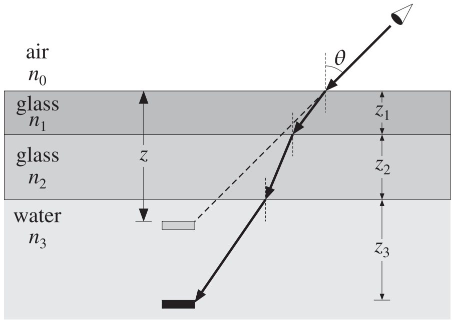
c. Suppose that the glass block is replaced by another dielectric with refractive index n. What is the minimum value of n in order that all entering rays will suffer total internal reflection at the top side?

7.9 An underwater object is viewed from air at an angle θ through a glass plate, as shown below. Let $z = z_1 + z_2$ be the actual depth of the object from the air surface, where z_1 is the thickness of the glass plate, and let n_1, n_2 be the refractive indices of the glass and water. Show that the apparent depth of the object is given by:

$$z' = \frac{z_1\cos\theta}{\sqrt{n_1^2 - \sin^2\theta}} + \frac{z_2\cos\theta}{\sqrt{n_2^2 - \sin^2\theta}}$$



7.10 An underwater object is viewed from air at an angle θ through two glass plates of refractive indices n_1, n_2 and thicknesses z_1, z_2, as shown below. Let z_3 be the depth of the object within the water.

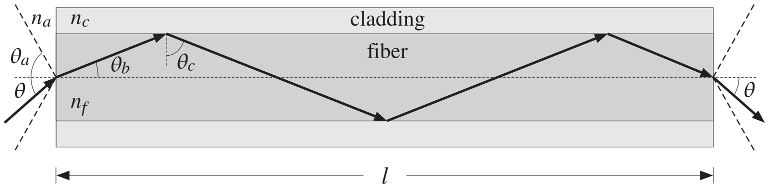


a. Express the apparent depth z of the object in terms of the quantities $\theta, n_0, n_1, n_2, n_3$ and z_1, z_2, z_3.

b. Generalize the results of the previous two problems to an arbitrary number of layers.

c. Consider also the continuous limit in which the body of water is inhomogeneous with a refractive index $n(z)$ given as a function of the depth z.

7.11 As shown below, light must be launched from air into an optical fiber at an angle $\theta \le \theta_a$ in order to propagate by total internal reflection.



a. Show that the acceptance angle is given by:

$$\sin\theta_a = \frac{\sqrt{n_f^2 - n_c^2}}{n_a}$$

b. For a fiber of length l, show that the exiting ray, at the opposite end, is exiting at the same angle θ as the incidence angle.

1. Show that the propagation delay time through this fiber, for a ray entering at an angle θ, is given as follows, where $t_0 = l/c_0$:

$$t(\theta) = \frac{t_0 n_f^2}{\sqrt{n_f^2 - n_a^2\sin^2\theta}}$$

c. What angles θ correspond to the maximum and minimum delay times? Show that the difference between the maximum and minimum delay times is given by:

$$\Delta t = t_{\max} - t_{\min} = \frac{t_0 n_f(n_f - n_c)}{n_c}$$

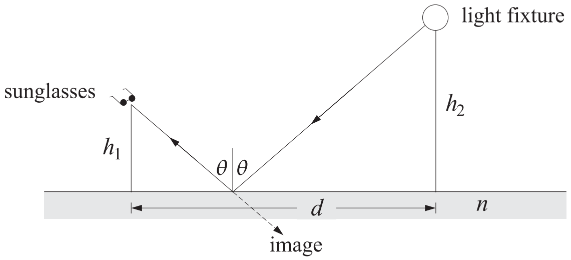
Such travel time delays cause "modal dispersion," that can limit the rate at which digital data may be transmitted (typically, the data rate must be $f_{\text{bps}} \le 1/(2\Delta t)$).

7.12 You are walking along the hallway in your classroom building wearing polaroid sunglasses and looking at the reflection of a light fixture on the waxed floor. Suddenly, at a distance d from the light fixture, the reflected image momentarily disappears. Show that the refractive index of the reflecting floor can be determined from the ratio of distances:

$$n = \frac{d}{h_1 + h_2}$$

where h_1 is your height and h_2 that of the light fixture. You may assume that light from the fixture is unpolarized, that is, a mixture of 50% TE and 50% TM, and that the polaroid sunglasses are designed to filter out horizontally polarized light. Explain your reasoning.[†]

[†] See, H. A. Smith, "Measuring Brewster's Angle Between Classes," *Physics Teacher*, Febr. 1979, p.109.



7.13 First, prove Eq. (7.12.13) from Eqs. (7.12.11). Then, show the following relationships among the angles $\theta, \theta_r, \theta'$:

$$\frac{\tan(\theta/2)}{\tan(\theta'/2)} = \sqrt{\frac{1-\beta}{1+\beta}}\,, \qquad \frac{\tan(\theta_r/2)}{\tan(\theta'/2)} = \sqrt{\frac{1+\beta}{1-\beta}}\,, \qquad \frac{\tan(\theta_r/2)}{\tan(\theta/2)} = \frac{1-\beta}{1+\beta}$$

7.14 A TM plane wave is incident obliquely on a moving interface as shown in Fig. 7.12.1. Show that the Doppler-shifted frequencies of the reflected and transmitted waves are still given by Eqs. (7.12.14) and (7.12.16). Moreover, show that the Brewster angle is given by:

$$\cos\theta_B = \frac{1+\beta\sqrt{n^2+1}}{\beta+\sqrt{n^2+1}}$$

8

Multilayer Film Applications

8.1 Multilayer Dielectric Structures at Oblique Incidence

Using the matching and propagation matrices for transverse fields that we discussed in Sec. 7.3, we derive here the layer recursions for multiple dielectric slabs at oblique incidence.

Fig. 8.1.1 shows such a multilayer structure. The layer recursions relate the various field quantities, such as the electric fields and the reflection responses, at the *left* of each interface.

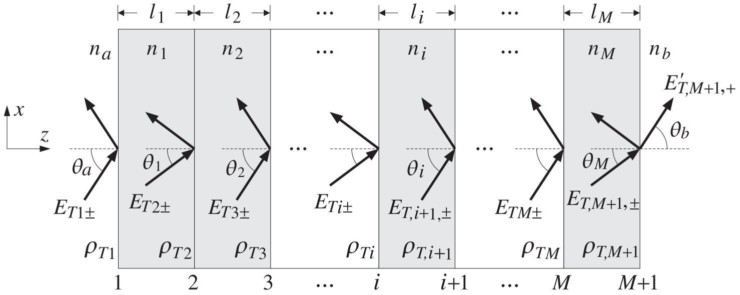


Fig. 8.1.1 Oblique incidence on multilayer dielectric structure.

We assume that there are no incident fields from the right side of the structure. The reflection/refraction angles in each medium are related to each other by Snel's law applied to each of the $M + 1$ interfaces:

$$\boxed{n_a \sin\theta_a = n_i \sin\theta_i = n_b \sin\theta_b}\,, \quad i = 1, 2, \dots, M \qquad (8.1.1)$$

It is convenient also to define by Eq. (7.3.8) the propagation phases or *phase thicknesses* for each of the M layers, that is, the quantities $\delta_i = k_{zi} l_i$. Using $k_{zi} = k_0 n_i \cos\theta_i$, where k_0 is the free-space wavenumber, $k_0 = \omega/c_0 = 2\pi f/c_0 = 2\pi/\lambda$, we have for $i = 1, 2, \dots, M$:

$$\boxed{\delta_i = \frac{\omega}{c_0}\,n_i l_i\cos\theta_i = \frac{2\pi}{\lambda}\,n_i l_i\cos\theta_i = \frac{2\pi}{\lambda}\,l_i n_i\sqrt{1-\frac{n_a^2\sin^2\theta_a}{n_i^2}}} \tag{8.1.2}$$

where we used Eq. (8.1.1) to write $\cos\theta_i = \sqrt{1-\sin^2\theta_i} = \sqrt{1-n_a^2\sin^2\theta_a/n_i^2}$. The transverse reflection coefficients at the $M+1$ interfaces are defined as in Eq. (6.1.1):

$$\rho_{Ti} = \frac{n_{T,i-1}-n_{Ti}}{n_{T,i-1}+n_{Ti}}\,,\quad i = 1,2,\dots,M+1 \tag{8.1.3}$$

where we set $n_{T0} = n_{Ta}$, as in Sec. 6.1. and $n_{T,M+1} = n_{Tb}$. The transverse refractive indices are defined in each medium by Eq. (7.2.13):

$$\boxed{n_{Ti} = \begin{cases} \dfrac{n_i}{\cos\theta_i}\,, & \text{TM polarization} \\ n_i\cos\theta_i\,, & \text{TE polarization}\end{cases}}\,,\qquad i = a,1,2,\dots,M,b \tag{8.1.4}$$

To obtain the layer recursions for the electric fields, we apply the propagation matrix (7.3.5) to the fields at the left of interface $i+1$ and propagate them to the right of the interface i, and then, apply a matching matrix (7.3.11) to pass to the left of that interface:

$$\begin{bmatrix} E_{Ti+} \\ E_{Ti-}\end{bmatrix} = \frac{1}{\tau_{Ti}}\begin{bmatrix} 1 & \rho_{Ti} \\ \rho_{Ti} & 1\end{bmatrix}\begin{bmatrix} e^{j\delta_i} & 0 \\ 0 & e^{-j\delta_i}\end{bmatrix}\begin{bmatrix} E_{T,i+1,+} \\ E_{T,i+1,-}\end{bmatrix}$$

Multiplying the matrix factors, we obtain:

$$\begin{bmatrix} E_{Ti+} \\ E_{Ti-}\end{bmatrix} = \frac{1}{\tau_{Ti}}\begin{bmatrix} e^{j\delta_i} & \rho_{Ti}e^{-j\delta_i} \\ \rho_{Ti}e^{j\delta_i} & e^{-j\delta_i}\end{bmatrix}\begin{bmatrix} E_{T,i+1,+} \\ E_{T,i+1,-}\end{bmatrix},\quad i = M, M-1,\dots,1 \tag{8.1.5}$$

This is identical to Eqs. (6.1.2) with the substitutions $k_i l_i \to \delta_i$ and $\rho_i \to \rho_{Ti}$. The recursion is initialized at the left of the $(M+1)$st interface by performing an additional matching to pass to the right of that interface:

$$\begin{bmatrix} E_{T,M+1,+} \\ E_{T,M+1,-}\end{bmatrix} = \frac{1}{\tau_{T,M+1}}\begin{bmatrix} 1 & \rho_{T,M+1} \\ \rho_{T,M+1} & 1\end{bmatrix}\begin{bmatrix} E'_{T.M+1,+} \\ 0\end{bmatrix} \tag{8.1.6}$$

It follows now from Eq. (8.1.5) that the reflection responses, $\Gamma_{Ti} = E_{Ti-}/E_{Ti+}$, will satisfy the identical recursions as Eq. (6.1.5):

$$\boxed{\Gamma_{Ti} = \frac{\rho_{Ti}+\Gamma_{T,i+1}e^{-2j\delta_i}}{1+\rho_{Ti}\Gamma_{T,i+1}e^{-2j\delta_i}}}\,,\quad i = M, M-1,\dots,1 \tag{8.1.7}$$

and initialized at $\Gamma_{T,M+1} = \rho_{T,M+1}$. Similarly, we obtain the following recursions for the total transverse electric and magnetic fields at each interface (they are continuous across each interface):

$$\begin{bmatrix} E_{Ti} \\ H_{Ti}\end{bmatrix} = \begin{bmatrix} \cos\delta_i & j\eta_{Ti}\sin\delta_i \\ j\eta_{Ti}^{-1}\sin\delta_i & \cos\delta_i\end{bmatrix}\begin{bmatrix} E_{T,i+1} \\ H_{T,i+1}\end{bmatrix},\quad i = M, M-1,\dots,1 \tag{8.1.8}$$

where η_{Ti} are the transverse characteristic impedances defined by Eq. (7.2.12) and related to the refractive indices by $\eta_{Ti} = \eta_0/n_{Ti}$. The wave impedances, $Z_{Ti} = E_{Ti}/H_{Ti}$, satisfy the following recursions initialized by $Z_{T,M+1} = \eta_{Tb}$:

$$\boxed{Z_{Ti} = \eta_{Ti}\,\frac{Z_{T,i+1}+j\eta_{Ti}\tan\delta_i}{\eta_{Ti}+jZ_{T,i+1}\tan\delta_i}}\,,\quad i = M, M-1,\dots,1 \tag{8.1.9}$$

The MATLAB function `multidiel` that was introduced in Sec. 6.1 can also be used in the oblique case with two extra input arguments: the incidence angle from the left and the polarization type, TE or TM. Its full usage is as follows:

```
[Gamma1,Z1] = multidiel(n,L,lambda,theta,pol);     % multilayer dielectric structure
```

where `theta` is the angle $\theta = \theta_a$ and `pol` is one of the strings `'te'` or `'tm'`. If the angle and polarization arguments are omitted, the function defaults to normal incidence for which TE and TM are the same. The other parameters have the same meaning as in Sec. 6.1.

In using this function, it is convenient to normalize the wavelength λ and the optical lengths $n_i l_i$ of the layers to some reference wavelength λ_0. The frequency f will be normalized to the corresponding reference frequency $f_0 = c_0/\lambda_0$.

Defining the normalized thicknesses $L_i = n_i l_i/\lambda_0$, so that $n_i l_i = L_i\lambda_0$, and noting that $\lambda_0/\lambda = f/f_0$, we may write the phase thicknesses (8.1.2) in the normalized form:

$$\boxed{\delta_i = 2\pi\,\frac{\lambda_0}{\lambda}\,L_i\cos\theta_i = 2\pi\,\frac{f}{f_0}\,L_i\cos\theta_i}\,,\qquad i = 1,2,\dots,M \tag{8.1.10}$$

Typically, but not necessarily, the L_i are chosen to be quarter-wavelength long at λ_0, that is, $L_i = 1/4$. This way the same multilayer design can be applied equally well at microwave or at optical frequencies. Once the wavelength scale λ_0 is chosen, the physical lengths of the layers l_i can be obtained from $l_i = L_i\lambda_0/n_i$.

8.2 Lossy Multilayer Structures

The `multidiel` function can be revised to handle lossy media. The reflection response of the multilayer structure is still computed from Eq. (8.1.7) but with some changes. In Sec. 7.7 we discussed the general case when either one or both of the incident and transmitted media are lossy.

In the notation of Fig. 8.1.1, we may assume that the incident medium n_a is lossless and all the other ones, n_i, $i = 1,2,\dots,M,b$, are lossy (and nonmagnetic). To implement `multidiel`, one needs to know the real and imaginary parts of n_i as functions of frequency, that is, $n_i(\omega) = n_{Ri}(\omega) - jn_{Ii}(\omega)$, or equivalently, the complex dielectric constants of the lossy media:

$$\begin{aligned} \epsilon_i(\omega) &= \epsilon_{Ri}(\omega) - j\epsilon_{Ii}(\omega)\,,\quad i = 1,2,\dots,M,b \\ n_i(\omega) &= \sqrt{\frac{\epsilon_i(\omega)}{\epsilon_0}} = \sqrt{\frac{\epsilon_{Ri}(\omega)-j\epsilon_{Ii}(\omega)}{\epsilon_0}} = n_{Ri}(\omega) - jn_{Ii}(\omega) \end{aligned} \tag{8.2.1}$$

Snel's law given in Eq. (8.1.1) remains valid, except now the angles θ_i and θ_b are complex valued because n_i, n_b are. One can still define the transverse refractive indices n_{Ti} through Eq. (8.1.4) using the complex-valued n_i, and $\cos\theta_i$ given by:

$$\cos\theta_i = \sqrt{1-\sin^2\theta_i} = \sqrt{1-\frac{n_a^2\sin^2\theta_a}{n_i^2}}\,, \quad i = a, 1, 2\dots, M, b \tag{8.2.2}$$

The reflection coefficients defined in Eq. (8.1.3) are equivalent to those given in Eq. (7.7.2) for the case of arbitrary incident and transmitted media.

The phase thicknesses δ_i now become complex-valued and are given by $\delta_i = k_{zi}l_i$, where k_{zi} is computed as follows. From Snel's law we have $k_{xi} = k_{xa} = \omega\sqrt{\mu_0\epsilon_0}n_a\sin\theta_a = k_0 n_a\sin\theta_a$, where $k_0 = \omega\sqrt{\mu_0\epsilon_0} = \omega/c_0$ is the free-space wave number. Then,

$$k_{zi} = \sqrt{\omega^2\mu_0\epsilon_i - k_{xi}^2} = \frac{\omega}{c_0}\sqrt{n_i^2 - n_a^2\sin^2\theta_a}\,, \quad i = a, 1, \dots, M, b \tag{8.2.3}$$

Thus, the complex phase thicknesses are given by:

$$\delta_i = k_{zi}l_i = \frac{\omega l_i}{c_0}\sqrt{n_i^2 - n_a^2\sin^2\theta_a}\,, \quad i = 1, 2, \dots, M \tag{8.2.4}$$

Writing $c_0 = f_0\lambda_0$ for some reference frequency and wavelength, we may re-express (8.2.4) in terms of the normalized frequency and normalized physical lengths:

$$\delta_i = k_{zi}l_i = 2\pi\frac{f}{f_0}\frac{l_i}{\lambda_0}\sqrt{n_i^2 - n_a^2\sin^2\theta_a}\,, \quad i = 1, 2, \dots, M \tag{8.2.5}$$

To summarize, given the complex $n_i(\omega)$ as in Eq. (8.2.1) at each desired value of ω, we calculate $\cos\theta_i$ from Eq. (8.2.2), n_{Ti} and ρ_{Ti} from Eqs. (8.1.4) and (8.1.3), and thicknesses δ_i from Eq. (8.2.5). Then, we use (8.1.7) to calculate the reflection response. The MATLAB function `multidiel2` implements these steps, with usage:

```
[Gamma1,Z1] = multidiel2(n,l,f,theta,pol);        % lossy multilayer structure
```

Once Γ_1 is determined, one may calculate the power entering each layer as well as the power lost within each layer. The time-averaged power per unit area entering the ith layer is the z-component of the Poynting vector, which is given in terms of the transverse E, H fields as follows:

$$\mathcal{P}_i = \frac{1}{2}\mathrm{Re}\big(E_{Ti}H_{Ti}^*\big)\,, \quad i = 1, 2, \dots, M \tag{8.2.6}$$

The power absorbed within the ith layer is equal to the difference of the power entering the layer and the power leaving it:

$$\mathcal{P}_i^{\text{loss}} = \mathcal{P}_i - \mathcal{P}_{i+1}\,, \quad i = 1, 2, \dots, M \tag{8.2.7}$$

The transverse fields can be calculated by inverting the recursion (8.1.8), that is,

$$\begin{bmatrix} E_{T,i+1} \\ H_{T,i+1} \end{bmatrix} = \begin{bmatrix} \cos\delta_i & -j\eta_{Ti}\sin\delta_i \\ -j\eta_{Ti}^{-1}\sin\delta_i & \cos\delta_i \end{bmatrix}\begin{bmatrix} E_{Ti} \\ H_{Ti} \end{bmatrix}, \quad i = 1, 2, \dots, M \tag{8.2.8}$$

The recursion is initialized with the fields E_{T1}, H_{T1} at the first interface. These can be calculated with the help of Γ_1:

$$\begin{aligned} E_{T1} &= E_{T1+} + E_{T1-} = E_{T1+}(1+\Gamma_1) \\ H_{T1} &= \frac{1}{\eta_{Ta}}(E_{T1+} - E_{T1-}) = \frac{1}{\eta_{Ta}}E_{T1+}(1-\Gamma_1) \end{aligned} \tag{8.2.9}$$

where $\eta_{Ta} = \eta_0/n_{Ta}$. The field E_{T1+} is the transverse component of the incident field. If we denote the total incident field by E_{in}, then E_{T1+} will be given by:

$$E_{T1+} = \begin{cases} E_{\text{in}}\,, & \text{TE case} \\ E_{\text{in}}\cos\theta_a\,, & \text{TM case} \end{cases} \tag{8.2.10}$$

The total incident power (along the direction of the incident wave vector), its z-component, and the power entering the first layer will be given as follows (in both the TE and TM cases):

$$\mathcal{P}_{\text{in}} = \frac{1}{2\eta_a}|E_{\text{in}}|^2\,, \quad \mathcal{P}_{\text{in},z} = \mathcal{P}_{\text{in}}\cos\theta_a\,, \quad \mathcal{P}_1 = \mathcal{P}_{\text{in},z}\big(1-|\Gamma_1|^2\big) \tag{8.2.11}$$

where $\eta_a = \eta_0/n_a$. Thus, one can start with $E_{\text{in}} = \sqrt{2\eta_a\mathcal{P}_{\text{in}}}$, if the incident power is known.

8.3 Single Dielectric Slab

Many features of oblique incidence on multilayer slabs can be clarified by studying the single-slab case, shown in Fig. 8.3.1. Assuming that the media to the left and right are the same, $n_a = n_b$, it follows that $\theta_b = \theta_a$ and also that $\rho_{T1} = -\rho_{T2}$. Moreover, Snel's law implies $n_a\sin\theta_a = n_1\sin\theta_1$.

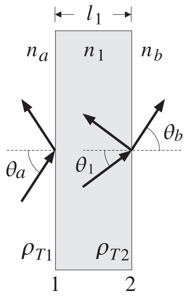


Fig. 8.3.1 Oblique incidence on single dielectric slab.

Because there are no incident fields from the right, the reflection response at the left of interface-2 is: $\Gamma_{T2} = \rho_{T2} = -\rho_{T1}$. It follows from Eq. (8.1.7) that the reflection response at the left of interface-1 will be:

$$\Gamma_{T1}=\frac{\rho_{T1}+\rho_{T2}e^{-2j\delta_1}}{1+\rho_{T1}\rho_{T2}e^{-2j\delta_1}}=\frac{\rho_{T1}(1-e^{-2j\delta_1})}{1-\rho_{T1}^2e^{-2j\delta_1}} \tag{8.3.1}$$

This is analogous to Eq. (5.5.4). According to Eq. (8.1.10), the phase thickness can be written in the following normalized form, where $L_1 = n_1 l_1/\lambda_0$:

$$\boxed{\delta_1 = 2\pi\frac{\lambda_0}{\lambda}L_1\cos\theta_1 = 2\pi\frac{f}{f_0}L_1\cos\theta_1 = \pi\frac{f}{f_1}} \tag{8.3.2}$$

$$\boxed{f_1=\frac{f_0}{2L_1\cos\theta_1}} \tag{8.3.3}$$

At frequencies that are integral multiples of f_1, $f = mf_1$, the reflection response vanishes because $2\delta_1 = 2\pi(mf_1)/f_1 = 2\pi m$ and $e^{-2j\delta_1}=1$. Similarly, at the half-integral multiples, $f=(m+0.5)f_1$, the response is maximum because $e^{-2j\delta_1}=-1$.

Because f_1 depends inversely on $\cos\theta_1$, then as the angle of incidence θ_a increases, $\cos\theta_1$ will decrease and f_1 will shift towards higher frequencies. The maximum shift will occur when θ_1 reaches its maximum refraction value $\theta_{1c} = \mathrm{asin}(n_a/n_1)$ (assuming $n_a < n_1$.)

Similar shifts occur for the 3-dB width of the reflection response notches. By the same calculation that led to Eq. (5.5.9), we find for the 3-dB width with respect to the variable δ_1:

$$\tan\left(\frac{\Delta\delta_1}{2}\right)=\frac{1-\rho_{T1}^2}{1+\rho_{T1}^2}$$

Setting $\Delta\delta_1 = \pi\Delta f/f_1$, we solve for the 3-dB width in frequency:

$$\Delta f = \frac{2f_1}{\pi}\,\mathrm{atan}\left(\frac{1-\rho_{T1}^2}{1+\rho_{T1}^2}\right) \tag{8.3.4}$$

The left/right bandedge frequencies are $f_1 \pm \Delta f/2$. The dependence of Δf on the incidence angle θ_a is more complicated here because ρ_{T1} also depends on it.

In fact, as θ_a tends to its grazing value $\theta_a \to 90^{\circ}$, the reflection coefficients for either polarization have the limit $|\rho_{T1}| \to 1$, resulting in zero bandwidth Δf. On the other hand, at the Brewster angle, $\theta_{aB} = \mathrm{atan}(n_1/n_a)$, the TM reflection coefficient vanishes, resulting in maximum bandwidth. Indeed, because $\mathrm{atan}(1) = \pi/4$, we have $\Delta f_{\max} = 2f_1\,\mathrm{atan}(1)/\pi = f_1/2$.

Fig. 8.3.2 illustrates some of these properties. The refractive indices were $n_a = n_b = 1$ and $n_1 = 1.5$. The optical length of the slab was taken to be half-wavelength at the reference wavelength λ_0, so that $n_1 l_1 = 0.5\lambda_0$, or, $L_1 = 0.5$.

The graphs show the TE and TM reflectances $|\Gamma_{T1}(f)|^2$ as functions of frequency for the angles of incidence $\theta_1 = 75^{\circ}$ and $\theta_a = 85^{\circ}$. The normal incidence case is also included for comparison.

The corresponding refracted angles were $\theta_1 = \mathrm{asin}\big(n_a\,\mathrm{asin}(\theta_a)/n_1\big) = 40.09^{\circ}$ and $\theta_1 = 41.62^{\circ}$. Note that the maximum refracted angle is $\theta_{1c} = 41.81^{\circ}$, and the Brewster angle, $\theta_{aB} = 56.31^{\circ}$.

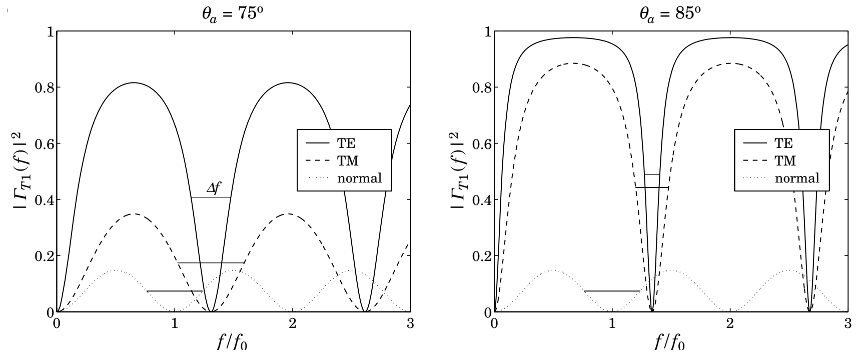


Fig. 8.3.2 TE and TM reflectances of half-wavelength slab.

The notch frequencies were $f_1 = f_0/(2L_1\cos\theta_1) = 1.31f_0$ and $f_1 = 1.34f_0$ for the angles $\theta_a = 75^{\circ}$ and 85°. At normal incidence we have $f_1 = f_0/(2L_1) = f_0$, because $L_1 = 0.5$.

The graphs also show the 3-dB widths of the notches, calculated from Eq. (8.3.4). The reflection responses were computed with the help of the function `multidiel` with the typical MATLAB code:

```
na = 1; nb = 1;
n1 = 1.5; L1 = 0.5;

f = linspace(0,3,401);
theta = 75;

G0 = abs(multidiel([na,n1,nb], L1, 1./f)).^2;
Ge = abs(multidiel([na,n1,nb], L1, 1./f, theta, 'te')).^2;
Gm = abs(multidiel([na,n1,nb], L1, 1./f, theta, 'tm')).^2;
```

The shifting of the notch frequencies and the narrowing of the notch widths is evident from the graphs. Had we chosen $\theta_a = \theta_{aB} = 56.31^{\circ}$, the TM response would have been identically zero because of the factor ρ_{T1} in Eq. (8.3.1).

The single-slab case is essentially a simplified version of a Fabry-Perot interferometer [598], used as a spectrum analyzer. At multiples of f_1, there are narrow transmittance bands. Because f_1 depends on $f_0/\cos\theta_1$, the interferometer serves to separate different frequencies f_0 in the input by mapping them onto different angles θ_1.

Next, we look at three further applications of the single-slab case: (a) frustrated total internal reflection, (b) surface plasmon resonance, and (c) the perfect lens property of negative-index media.

8.4 Frustrated Total Internal Reflection

As we discussed in Sec. 7.5, when a wave is incident at an angle greater than the total internal reflection (TIR) angle from an optically denser medium n_a onto a rarer medium

n_b, with $n_a > n_b$, then there is 100 percent reflection. The transmitted field into the rarer medium n_b is evanescent, decaying exponentially with distance.

However, if an object or another medium is brought near the interface from the n_b side, the evanescent field is “frustrated” and can couple into a propagating wave. For example, if another semi-infinite medium n_a is brought close to the interface, then the evanescent field can “tunnel” through to the other side, emerging as an attenuated version of the incident wave. This effect is referred to as “frustrated” total internal reflection.

Fig. 8.4.1 shows how this may be realized with two 45° prisms separated by a small air gap. With $n_a = 1.5$ and $n_b = 1$, the TIR angle is $\theta_c = \operatorname{asin}(n_b/n_a) = 41.8^\circ$, therefore, $\theta = 45^\circ > \theta_c$. The transmitted fields into the air gap reach the next prism with an attenuated magnitude and get refracted into a propagating wave that emerges at the same angle θ.

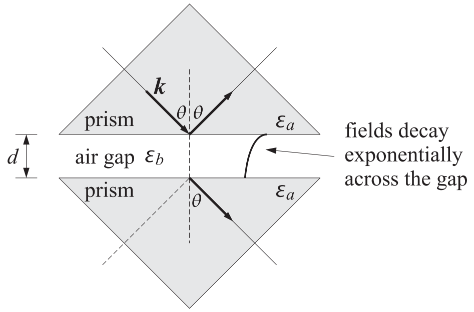


Fig. 8.4.1 Frustrated total internal reflection between two prisms separated by an air gap.

Fig. 8.4.2 shows an equivalent problem of two identical semi-infinite media n_a, separated by a medium n_b of length d. Let $\varepsilon_a = n_a^2$, $\varepsilon_b = n_b^2$ be the relative dielectric constants. The components of the wavevectors in media n_a and n_b are:

$$\begin{aligned}
&k_x = k_0 n_a \sin\theta\,, \quad k_0 = \frac{\omega}{c_0}\\
&k_{za} = \sqrt{k_0^2 n_a^2 - k_x^2} = k_0 n_a \cos\theta\\
&k_{zb} = \begin{cases} k_0\sqrt{n_b^2 - n_a^2\sin^2\theta}\,, & \text{if} \quad \theta \le \theta_c \\ -jk_0\sqrt{n_a^2\sin^2\theta - n_b^2} = -j\alpha_{zb}\,, & \text{if} \quad \theta \ge \theta_c \end{cases}
\end{aligned} \tag{8.4.1}$$

where $\sin\theta_c = n_b/n_a$. Because of Snel's law, the k_x component is preserved across the interfaces. If $\theta > \theta_c$, then k_{zb} is pure imaginary, that is, evanescent.
The transverse reflection and transmission responses are:

$$\begin{aligned}
\Gamma &= \frac{\rho_a + \rho_b e^{-2jk_{zb}d}}{1 + \rho_a\rho_b e^{-2jk_{zb}d}} = \frac{\rho_a(1 - e^{-2jk_{zb}d})}{1 - \rho_a^2 e^{-2jk_{zb}d}}\\
T &= \frac{\tau_a\tau_b e^{-jk_{zb}d}}{1 + \rho_a\rho_b e^{-2jk_{zb}d}} = \frac{(1 - \rho_a^2)e^{-jk_{zb}d}}{1 - \rho_a^2 e^{-2jk_{zb}d}}
\end{aligned} \tag{8.4.2}$$

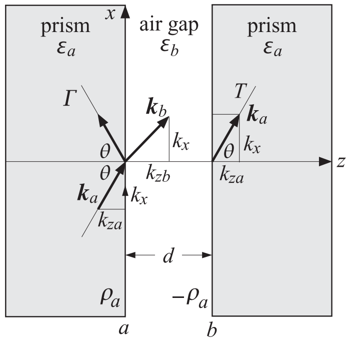


Fig. 8.4.2 Frustrated total internal reflection.

where ρ_a, ρ_b are the transverse reflection coefficients at the a, b interfaces and $\tau_a = 1 + \rho_a$ and $\tau_b = 1 + \rho_b$ are the transmission coefficients, and we used the fact that $\rho_b = -\rho_a$ because the media to the left and right of the slab are the same. For the two polarizations, ρ_a is given in terms of the above wavevector components as follows:

$$\rho_a^{\mathrm{TE}} = \frac{k_{za} - k_{zb}}{k_{za} + k_{zb}}\,, \quad \rho_a^{\mathrm{TM}} = \frac{k_{zb}\varepsilon_a - k_{za}\varepsilon_b}{k_{zb}\varepsilon_a + k_{za}\varepsilon_b} \tag{8.4.3}$$

For $\theta \le \theta_c$, the coefficient ρ_a is real-valued, and for $\theta \ge \theta_c$, it is unimodular, $|\rho_a| = 1$, given explicitly by

$$\rho_a^{\mathrm{TE}} = \frac{n_a\cos\theta + j\sqrt{n_a^2\sin^2\theta - n_b^2}}{n_a\cos\theta - j\sqrt{n_a^2\sin^2\theta - n_b^2}}\,, \quad \rho_a^{\mathrm{TM}} = \frac{-jn_a\sqrt{n_a^2\sin^2\theta - n_b^2} - n_b^2\cos\theta}{-jn_a\sqrt{n_a^2\sin^2\theta + n_b^2} + n_b^2\cos\theta} \tag{8.4.4}$$

For all angles, it can be shown that $1 - |\Gamma|^2 = |T|^2$, which represents the amount of power the enters perpendicularly into interface a and exits from interface b. For the TIR case, Γ, T simplify into:

$$\Gamma = \frac{\rho_a(1 - e^{-2\alpha_{zb}d})}{1 - \rho_a^2 e^{-2\alpha_{zb}d}}\,, \quad T = \frac{(1 - \rho_a^2)e^{-\alpha_{zb}d}}{1 - \rho_a^2 e^{-2\alpha_{zb}d}}\,, \quad \alpha_{zb} = \frac{2\pi d}{\lambda_0}\sqrt{n_a^2\sin^2\theta - n_b^2} \tag{8.4.5}$$

where we defined the free-space wavelength through $k_0 = 2\pi/\lambda_0$. Setting $\rho_a = e^{j\phi_a}$, the magnitude responses are given by:

$$|\Gamma|^2 = \frac{\sinh^2(\alpha_{zb}d)}{\sinh^2(\alpha_{zb}d) + \sin^2\phi_a}\,, \quad |T|^2 = \frac{\sin^2\phi_a}{\sinh^2(\alpha_{zb}d) + \sin^2\phi_a} \tag{8.4.6}$$

For a prism with $n_a = 1.5$ and an air gap $n_b = 1$, Fig. 8.4.3 shows a plot of Eqs. (8.4.5) versus the distance d at the incidence angle $\theta = 45^\circ$. The reflectance becomes almost 100 percent for thickness of a few wavelengths.

Fig. 8.4.4 shows the reflectance versus angle over $0 \le \theta \le 90^\circ$ for the thicknesses $d = 0.4\lambda_0$ and $d = 0.5\lambda_0$. The TM reflection response vanishes at the Brewster angle $\theta_B = \operatorname{atan}(n_b/n_a) = 33.69^\circ$.

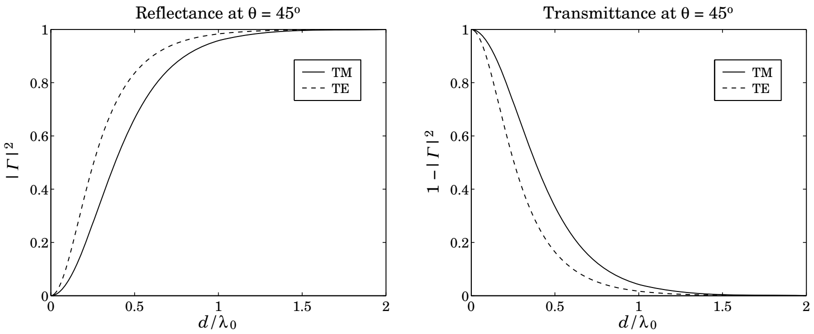


Fig. 8.4.3 Reflectance and transmittance versus thickness d.

The case $d = 0.5\lambda_0$ was chosen because the slab becomes a half-wavelength slab at normal incidence, that is, $k_{zb}d = 2\pi/2$ at $\theta = 0^\circ$, resulting in the vanishing of Γ as can be seen from Eq. (8.4.2).

The half-wavelength condition, and the corresponding vanishing of Γ, can be required at any desired angle $\theta_0 < \theta_c$, by demanding that $k_{zb}d = 2\pi/2$ at that angle, which fixes the separation d:

$$k_{zb}d = \pi \quad \Rightarrow \quad \frac{2\pi d}{\lambda_0}\sqrt{n_b^2 - n_a^2\sin^2\theta_0} = \pi \quad \Rightarrow \quad d = \frac{\lambda_0}{2\sqrt{n_b^2 - n_a^2\sin^2\theta_0}}$$

Fig. 8.4.5 depicts the case $\theta_0 = 20^\circ$, which fixes the separation to be $d = 0.5825\lambda_0$.

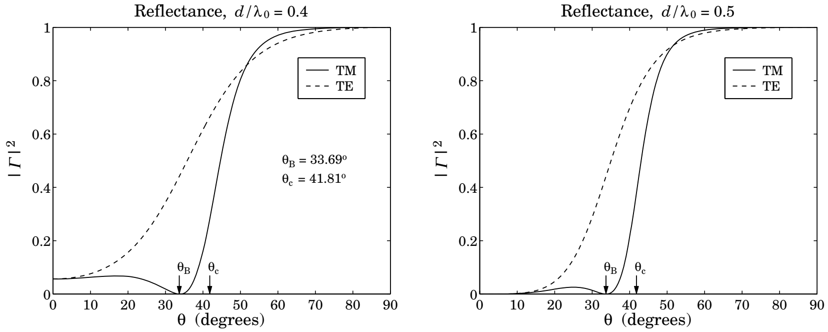


Fig. 8.4.4 Reflectance versus angle of incidence.

The fields within the air gap can be determined using the layer recursions (8.1.5). Let E_{a+} be the incident transverse field at the left side of the interface a, and $E_\pm$ the transverse fields at the right side. Using Eq. (8.1.5) and (8.1.6), we find for the TIR case:

$$E_+ = \frac{(1+\rho_a)E_{a+}}{1-\rho_a^2 e^{-2\alpha_{zb}d}}\,, \quad E_- = \frac{-\rho_a e^{-2\alpha_{zb}d}(1+\rho_a)E_{a+}}{1-\rho_a^2 e^{-2\alpha_{zb}d}} \tag{8.4.7}$$

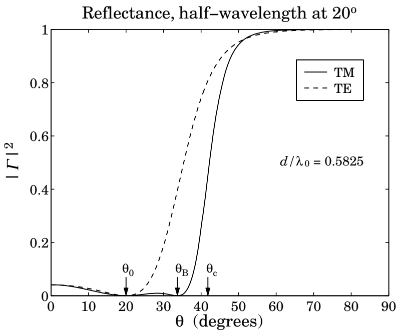


Fig. 8.4.5 Reflectance vanishes at $\theta_0 = 20^\circ$.

The transverse electric field within the air gap will be then $E_T(z) = E_+e^{-\alpha_{zb}z} + E_-e^{\alpha_{zb}z}$, and similarly for the magnetic field. Using (8.4.7) we find:

$$\begin{aligned} E_T(z) &= \left[\frac{1+\rho_a}{1-\rho_a^2 e^{-2\alpha_{zb}d}}\right]\left[e^{-\alpha_{zb}z} - \rho_a e^{-2\alpha_{zb}d}e^{\alpha_{zb}z}\right]E_{a+} \\ H_T(z) &= \left[\frac{1-\rho_a}{1-\rho_a^2 e^{-2\alpha_{zb}d}}\right]\left[e^{-\alpha_{zb}z} + \rho_a e^{-2\alpha_{zb}d}e^{\alpha_{zb}z}\right]\frac{E_{a+}}{\eta_{aT}} \end{aligned} \tag{8.4.8}$$

where η_{aT} is the transverse impedance of medium n_a, that is, with $\eta_a = \eta_0/n_a$:

$$\eta_{aT} = \begin{cases} \eta_a\cos\theta_a\,, & \text{TM, or parallel polarization} \\ \eta_a/\cos\theta_a\,, & \text{TE, or perpendicular polarization} \end{cases}$$

It is straightforward to verify that the transfer of power across the gap is independent of the distance z and given by

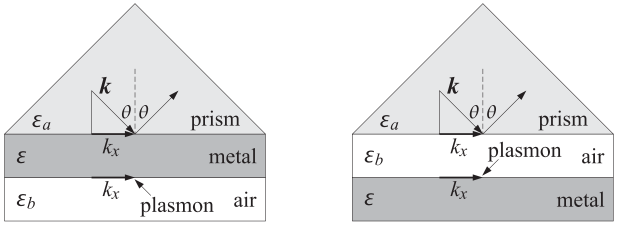
$$\mathcal{P}_z(z) = \frac{1}{2}\,\mathrm{Re}\big[E_T(z)H_T^*(z)\big] = \big(1-|\Gamma|^2\big)\,\frac{|E_{a+}|^2}{2\eta_{aT}}$$

Frustrated total internal reflection has several applications [516–552], such as internal reflection spectroscopy, sensors, fingerprint identification, surface plasmon resonance, and high resolution microscopy. In many of these applications, the air gap is replaced by another, possibly lossy, medium. The above formulation remains valid with the replacement $\varepsilon_b = n_b^2 \to \varepsilon_b = \varepsilon_{br} - j\varepsilon_{bi}$, where the imaginary part ε_{ri} characterizes the losses.

8.5 *Surface Plasmon Resonance*

We saw in Sec. 7.7 that surface plasmons are TM waves that can exist at an interface between air and metal, and that their wavenumber k_x of propagation along the interface is larger that its free-space value at the same frequency. Therefore, such plasmons cannot couple directly to plane waves incident on the interface.

However, if the incident TM plane wave is from a dielectric and from an angle that is greater than the angle of total internal reflection, then the corresponding wavenumber will be greater than its vacuum value and it could excite a plasmon wave along the interface. Fig. 8.5.1 depicts two possible configurations of how this can be accomplished.



Kretschmann-Raether Otto

Fig. 8.5.1 Kretschmann-Raether and Otto configurations.

In the so-called Kretschmann-Raether configuration [555,558], a thin metal film of thickness of a fraction of a wavelength is sandwiched between a prism and air and the incident wave is from the prism side. In the Otto configuration [556], there is an air gap between the prism and the metal. The two cases are similar, but we will consider in greater detail the Kretschmann-Raether configuration, which is depicted in more detail in Fig. 8.5.2.

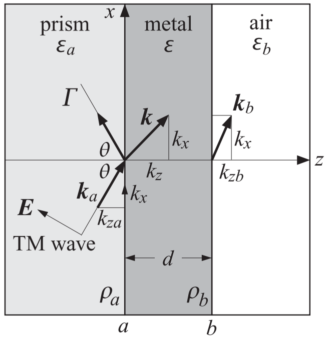


Fig. 8.5.2 Surface plasmon resonance excitation by total internal reflection.

The relative dielectric constant ε_a and refractive index n_a of the prism are related by $\varepsilon_a = n_a^2$. The air side has $\varepsilon_b = n_b^2 = 1$, but any other lossless dielectric will do as long as it satisfies $n_b < n_a$. The TIR angle is $\sin\theta_c = n_b/n_a$, and the angle of incidence from the prism side is assumed to be $\theta \geq \theta_c$ so that

$$k_x = k_0 n_a \sin\theta \geq k_0 n_b\,, \quad k_0 = \frac{\omega}{c_0} \tag{8.5.1}$$

Because of Snel's law, the k_x component of the wavevector along the interface is preserved across the media. The z-components in the prism and air sides are given by:

$$\begin{aligned} k_{za} &= \sqrt{k_0^2 n_a^2 - k_x^2} = k_0 n_a \cos\theta \\ k_{zb} &= -j\alpha_{zb} = -j\sqrt{k_x^2 - k_0^2 n_b^2} = -jk_0\sqrt{n_a^2 \sin^2\theta - n_b^2} \end{aligned} \tag{8.5.2}$$

where k_{zb} is pure imaginary because of the TIR assumption. Therefore, the transmitted wave into the ε_b medium attenuates exponentially like $e^{-jk_{zb}z} = e^{-\alpha_{zb}z}$.

For the metal layer, we assume that its relative dielectric constant is $\varepsilon = -\varepsilon_r - j\varepsilon_i$, with a negative real part ($\varepsilon_r > 0$) and a small negative imaginary part ($0 < \varepsilon_i \ll \varepsilon_r$) that represents losses. Moreover, in order for a surface plasmon wave to be supported on the ε-ε_b interface, we must further assume that $\varepsilon_r > \varepsilon_b$. The k_z component within the metal will be complex-valued with a dominant imaginary part:

$$k_z = -j\sqrt{k_x^2 - k_0^2\varepsilon} = -j\sqrt{k_x^2 + k_0^2(\varepsilon_r + j\varepsilon_i)} = -jk_0\sqrt{n_a^2\sin^2\theta + \varepsilon_r + j\varepsilon_i} \tag{8.5.3}$$

If there is a surface plasmon wave on the ε-ε_b interface, then as we saw in Sec. 7.7, it will be characterized by the specific values of k_x, k_z, k_{zb}:

$$k_{x0} = \beta_{x0} - j\alpha_{x0} = k_0\sqrt{\frac{\varepsilon\varepsilon_b}{\varepsilon + \varepsilon_b}}\,, \quad k_{z0} = -\frac{k_0\varepsilon}{\sqrt{\varepsilon + \varepsilon_b}}\,, \quad k_{zb0} = \frac{k_0\varepsilon_b}{\sqrt{\varepsilon + \varepsilon_b}} \tag{8.5.4}$$

Using Eq. (7.11.10), we have approximately to lowest order in ε_i:

$$\beta_{x0} = k_0\sqrt{\frac{\varepsilon_r\varepsilon_b}{\varepsilon_r - \varepsilon_b}}\,, \quad \alpha_{x0} = k_0\left(\frac{\varepsilon_r\varepsilon_b}{\varepsilon_r - \varepsilon_b}\right)^{3/2}\frac{\varepsilon_i}{2\varepsilon_r^2} \tag{8.5.5}$$

and similarly for k_{z0}, which has a small real part and a dominant imaginary part:

$$k_{z0} = \beta_{z0} - j\alpha_{z0}\,, \quad \alpha_{z0} = \frac{k_0\varepsilon_r}{\sqrt{\varepsilon_r - \varepsilon_b}}\,, \quad \beta_{z0} = \frac{k_0(\varepsilon_r - 2\varepsilon_b)\varepsilon_i}{(\varepsilon_r - \varepsilon_b)^{3/2}} \tag{8.5.6}$$

If the incidence angle θ is such that k_x is near the real-part of k_{x0}, that is, $k_x = k_0 n_a \sin\theta = \beta_{x0}$, then a resonance takes place exciting the surface plasmon wave. Because of the finite thickness d of the metal layer and the assumed losses ε_i, the actual resonance condition is not $k_x = \beta_{x0}$, but is modified by a small shift: $k_x = \beta_{x0} + \tilde{\beta}_{x0}$, to be determined shortly.

At the resonance angle there is a sharp drop of the reflection response measured at the prism side. Let ρ_a, ρ_b denote the TM reflection coefficients at the ε_a-ε and ε-ε_b interfaces, as shown in Fig. 8.5.2. The corresponding TM reflection response of the structure will be given by:

$$\Gamma = \frac{\rho_a + \rho_b e^{-2jk_z d}}{1 + \rho_a\rho_b e^{-2jk_z d}} = \frac{\rho_a + \rho_b e^{-2\alpha_z d}e^{-2j\beta_z d}}{1 + \rho_a\rho_b e^{-2\alpha_z d}e^{-2j\beta_z d}} \tag{8.5.7}$$

where d is the thickness of the metal layer and $k_z = \beta_z - j\alpha_z$ is given by Eq. (8.5.3). The TM reflection coefficients are given by:

$$\rho_a = \frac{k_z\varepsilon_a - k_{za}\varepsilon}{k_z\varepsilon_a + k_{za}\varepsilon}\,, \quad \rho_b = \frac{k_{zb}\varepsilon - k_z\varepsilon_b}{k_{zb}\varepsilon + k_z\varepsilon_b} \tag{8.5.8}$$

where k_{za}, k_{zb} are given by (8.5.2). Explicitly, we have for $\theta \geq \theta_c$:

$$\begin{aligned}
\rho_a &= -\frac{\varepsilon\sqrt{k_0^2\varepsilon_a - k_x^2} + j\varepsilon_a\sqrt{k_x^2 - k_0^2\varepsilon}}{\varepsilon\sqrt{k_0^2\varepsilon_a - k_x^2} - j\varepsilon_a\sqrt{k_x^2 - k_0^2\varepsilon}} = -\frac{\varepsilon\cos\theta + jn_a\sqrt{\varepsilon_a\sin^2\theta - \varepsilon}}{\varepsilon\cos\theta - jn_a\sqrt{\varepsilon_a\sin^2\theta - \varepsilon}} \\
\rho_b &= \frac{\varepsilon\sqrt{k_x^2 - k_0^2\varepsilon_b} - \varepsilon_b\sqrt{k_x^2 - k_0^2\varepsilon}}{\varepsilon\sqrt{k_x^2 - k_0^2\varepsilon_b} + \varepsilon_b\sqrt{k_x^2 - k_0^2\varepsilon}} = \frac{\varepsilon\sqrt{\varepsilon_a\sin^2\theta - \varepsilon_b} - \varepsilon_b\sqrt{\varepsilon_a\sin^2\theta - \varepsilon}}{\varepsilon\sqrt{\varepsilon_a\sin^2\theta - \varepsilon_b} + \varepsilon_b\sqrt{\varepsilon_a\sin^2\theta - \varepsilon}}
\end{aligned} \tag{8.5.9}$$

We note that for the plasmon resonance to be excited through such a configuration, the metal must be assumed to be slightly lossy, that is, $\varepsilon_i \neq 0$. If we assume that it is lossless with a negative real part, $\varepsilon = -\varepsilon_r$, then, ρ_a becomes a unimodular complex number, $|\rho_a| = 1$, for all angles θ, while ρ_b remains real-valued for $\theta \geq \theta_c$, and also k_z is pure imaginary, $\beta_z = 0$. Hence, it follows that:

$$|\Gamma|^2 = \frac{|\rho_a|^2 + 2\,\mathrm{Re}(\rho_a)\,\rho_b e^{-2\alpha_z d} + \rho_b^2 e^{-4\alpha_z d}}{1 + 2\,\mathrm{Re}(\rho_a)\,\rho_b e^{-2\alpha_z d} + |\rho_a|^2\rho_b^2 e^{-4\alpha_z d}} = 1$$

Thus, it remains flat for $\theta \geq \theta_c$. For $\theta \leq \theta_c$, ρ_a is still unimodular, and ρ_b also becomes unimodular, $|\rho_b| = 1$. Setting $\rho_a = e^{j\phi_a}$ and $\rho_b = e^{j\phi_b}$, we find for $\theta \leq \theta_c$:

$$|\Gamma|^2 = \left|\frac{e^{j\phi_a} + e^{j\phi_b}e^{-2\alpha_z d}}{1 + e^{j\phi_a}e^{j\phi_b}e^{-2\alpha_z d}}\right|^2 = \frac{1 + 2\cos(\phi_a - \phi_b)\,e^{-2\alpha_z d} + e^{-4\alpha_z d}}{1 + 2\cos(\phi_a + \phi_b)\,e^{-2\alpha_z d} + e^{-4\alpha_z d}} \tag{8.5.10}$$

which remains almost flat, exhibiting a slight variation with the angle for $\theta \leq \theta_c$.

As an example, consider a quartz prism with $n_a = 1.5$, coated with a silver film of thickness of $d = 50$ nm, and air on the other side $\varepsilon_b = 1$. The relative refractive index of the metal is taken to be $\varepsilon = -16 - 0.5j$ at the free-space wavelength of $\lambda_0 = 632$ nm. The corresponding free-space wave number is $k_0 = 2\pi/\lambda_0 = 9.94\ \mu$m.

Fig. 8.5.3 shows the TM reflection response (8.5.7) versus angle. The TIR angle is $\theta_c = \mathrm{asin}(n_b/n_a) = 41.81^{\mathrm{o}}$. The plasmon resonance occurs at the angle $\theta_{\mathrm{res}} = 43.58^{\mathrm{o}}$. The graph on the right shows an expanded view over the angle range $41^{\mathrm{o}} \leq \theta \leq 45^{\mathrm{o}}$. Both angles θ_c and θ_{res} are indicated on the graphs as black dots.

The computation can be carried out with the help of the MATLAB function `multidiel1.m`, or alternatively `multidiel.m`, with the sample code:

```
na = 1.5;  ea = na^2;                          % prism side
er = 16;   ei = 0.5; ep = -er-j*ei;            % silver layer
nb = 1;    eb = nb^2;                          % air side
d = 50; la0 = 632;                             % in units of nanometers

th = linspace(0,89,8901);                      % incident angle in degrees

n1 = sqrte(ep);                                % evanescent SQRT, needed if εi = 0
L1 = n1*d/la0;                                 % complex optical length in units of λ0
n = [na, n1, nb];                              % input to multidiel1


for i=1:length(th),                                         % TM reflectance
   Ga(i) = abs(multidiel1(n, L1, 1, th(i), 'tm')).^2;      % at λ/λ0 = 1
end

plot(th,Ga);
```

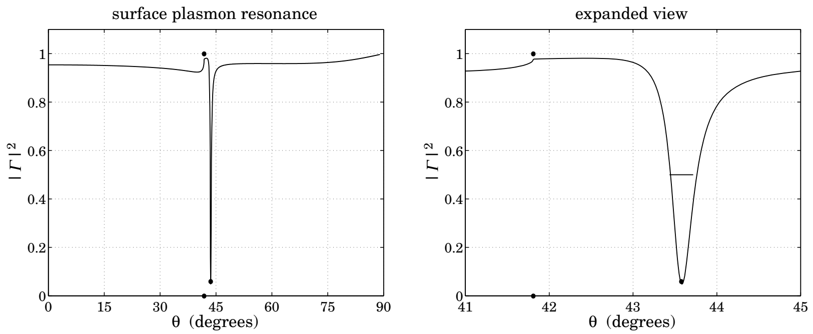


Fig. 8.5.3 Surface plasmon resonance.

Fig. 8.5.4 shows the reflection response when the metal is assumed to be lossless with $\varepsilon = -16$, all the other parameters being the same. As expected, there is no resonance and the reflectance stays flat for $\theta \geq \theta_c$, with mild variation for $\theta < \theta_c$.

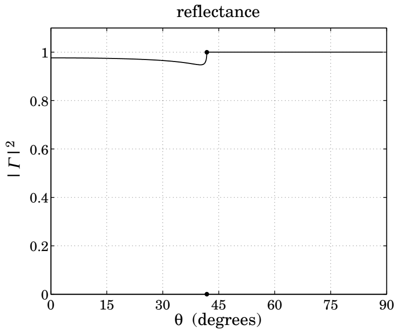


Fig. 8.5.4 Absence of resonance when metal is assumed to be lossless.

Let E_{a+}, E_{a-} be the forward and backward transverse electric fields at the left side of interface a. The fields at the right side of the interface can be obtained by inverting the matching matrix:

$$\begin{bmatrix} E_{a+} \\ E_{a-} \end{bmatrix} = \frac{1}{1+\rho_a}\begin{bmatrix} 1 & \rho_a \\ \rho_a & 1 \end{bmatrix}\begin{bmatrix} E_+ \\ E_- \end{bmatrix} \quad\Rightarrow\quad \begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \frac{1}{1-\rho_a}\begin{bmatrix} 1 & -\rho_a \\ -\rho_a & 1 \end{bmatrix}\begin{bmatrix} E_{a+} \\ E_{a-} \end{bmatrix}$$

Setting $E_{a-} = \Gamma E_{a+}$, with Γ given by Eq. (8.5.7), we obtain:

$$\begin{aligned}
E_+ &= \frac{1 - \rho_a\Gamma}{1 - \rho_a}\,E_{a+} = \frac{(1+\rho_a)\,E_{a+}}{1 + \rho_a\rho_b e^{-2jk_z d}} \\
E_+ &= \frac{-\rho_a + \Gamma}{1 - \rho_a}\,E_{a+} = \frac{\rho_b e^{-2jk_z d}(1+\rho_a)\,E_{a+}}{1 + \rho_a\rho_b e^{-2jk_z d}}
\end{aligned}$$

The transverse electric and magnetic fields within the metal layer will be given by:

$$E_T(z) = E_+ e^{-jk_z z} + E_- e^{jk_z z}\,, \quad H_T(z) = \frac{1}{\eta_T}\big[E_+ e^{-jk_z z} - E_- e^{jk_z z}\big]$$

Using the relationship $\eta_T/\eta_{aT} = (1+\rho_a)/(1-\rho_a)$, we have:

$$\begin{aligned} E_T(z) &= \left[\frac{1+\rho_a}{1+\rho_a\rho_b e^{-2jk_z d}}\right]\big[e^{-jk_z z} + \rho_b e^{-2jk_z d} e^{jk_z z}\big]E_{a+} \\ H_T(z) &= \left[\frac{1-\rho_a}{1+\rho_a\rho_b e^{-2jk_z d}}\right]\big[e^{-jk_z z} - \rho_b e^{-2jk_z d} e^{jk_z z}\big]\frac{E_{a+}}{\eta_{aT}} \end{aligned} \tag{8.5.11}$$

where $\eta_{aT} = \eta_a \cos\theta$ is the TM characteristic impedance of the prism. The power flow within the metal strip is described by the z-component of the Poynting vector:

$$\mathcal{P}(z) = \frac{1}{2}\,\text{Re}\big[E_T(z)H_T^*(z)\big] \tag{8.5.12}$$

The power entering the conductor at interface a is:

$$\mathcal{P}_{\text{in}} = (1 - |\Gamma|^2)\frac{|E_{a+}|^2}{2\eta_{aT}} = \frac{1}{2}\,\text{Re}\big[E_T(z)H_T^*(z)\big]\Big|_{z=0} \tag{8.5.13}$$

Fig. 8.5.5 shows a plot of the quantity $\mathcal{P}(z)/\mathcal{P}_{\text{in}}$ versus distance within the metal, $0 \le z \le d$, at the resonant angle of incidence $\theta = \theta_{\text{res}}$. Because the fields are evanescent in the right medium n_b, the power vanishes at interface b, that is, at $z = d$. The reflectance at the resonance angle is $|\Gamma|^2 = 0.05$, and therefore, the fraction of the incident power that enters the metal layer and is absorbed by it is $1 - |\Gamma|^2 = 0.95$.

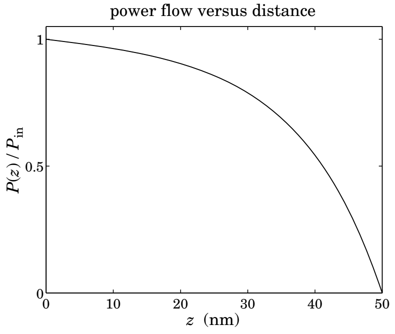


Fig. 8.5.5 Power flow within metal layer at the resonance angle $\theta_{\text{res}} = 43.58^{\text{o}}$.

The angle width of the resonance of Fig. 8.5.3, measured at the 3-dB level $|\Gamma|^2 = 1/2$, is very narrow, $\Delta\theta = 0.282^{\text{o}}$. The width $\Delta\theta$, as well as the resonance angle θ_{res}, and the optimum metal film thickness d, can be estimated by the following approximate procedure.

To understand the resonance property, we look at the behavior of Γ in the neighborhood of the plasmon wavenumber $k_x = k_{x0}$ given by (8.5.4). At this value, the TM reflection coefficient at the ε-ε_b interface develops a pole, $\rho_b = \infty$, which is equivalent to the condition $k_{zb0}\varepsilon + k_{z0}\varepsilon_b = 0$, with k_{zb0}, k_{z0} defined by Eq. (8.5.4).

In the neighborhood of this pole, $k_x \simeq k_{x0}$, ρ_b will be given by $\rho_b \simeq K_0/(k_x - k_{x0})$, where K_0 is the residue of the pole. It can be determined by:

$$K_0 = \lim_{k_x\to k_{x0}}(k_x - k_{x0})\rho_b = \lim_{k_x\to k_{x0}}(k_x - k_{x0})\frac{k_{zb}\varepsilon - k_z\varepsilon_b}{k_{zb}\varepsilon + k_z\varepsilon_b} = \frac{k_{zb}\varepsilon - k_z\varepsilon_b}{\dfrac{d}{dk_x}(k_{zb}\varepsilon + k_z\varepsilon_b)}\Bigg|_{k_x=k_{x0}}$$

The derivative dk_z/dk_x can be determined by differentiating $k_z^2 + k_x^2 = k_0^2\varepsilon$, that is, $k_z\,dk_z + k_x\,dk_x = 0$, which gives $dk_z/dk_x = -k_x/k_z$, and similarly for dk_{zb}/dk_x. It follows that:

$$K_0 = \frac{k_{zb0}\varepsilon - k_{z0}\varepsilon_b}{-\dfrac{k_{x0}}{k_{zb0}}\,\varepsilon - \dfrac{k_{x0}}{k_{z0}}\,\varepsilon_b}$$

Inserting k_{x0}, k_{z0}, k_{zb0} from Eq. (8.5.4), we obtain:

$$K_0 = k_0\left(\frac{2}{\varepsilon_b - \varepsilon}\right)\left(\frac{\varepsilon\varepsilon_b}{\varepsilon + \varepsilon_b}\right)^{3/2} \tag{8.5.14}$$

The reflection response can then be approximated near $k_x \simeq k_{x0}$ by

$$\Gamma \simeq \frac{\rho_a + \dfrac{K_0}{k_x - k_{x0}}\,e^{-2jk_z d}}{1 + \rho_a\dfrac{K_0}{k_x - k_{x0}}\,e^{-2jk_z d}}$$

The quantities ρ_a and $e^{-2jk_z d}$ can also be replaced by their values at k_{x0}, k_{z0}, k_{zb0}, thus obtaining:

$$\Gamma = \rho_{a0}\,\frac{k_x - k_{x0} + \rho_{a0}^{-1}K_0 e^{-2jk_{z0}d}}{k_x - k_{x0} + \rho_{a0}K_0 e^{-2jk_{z0}d}} \tag{8.5.15}$$

where

$$\rho_{a0} = \frac{k_{z0}\varepsilon_a - k_{za0}\varepsilon}{k_{z0}\varepsilon_a + k_{za0}\varepsilon} = \frac{\varepsilon_a + \sqrt{\varepsilon(\varepsilon_a - \varepsilon_b) + \varepsilon_a\varepsilon_b}}{\varepsilon_a - \sqrt{\varepsilon(\varepsilon_a - \varepsilon_b) + \varepsilon_a\varepsilon_b}}$$

which was obtained using $k_{za0} = \sqrt{k_0^2\varepsilon_a - k_{x0}^2}$ and Eqs. (8.5.4). Replacing $\varepsilon = -\varepsilon_r - j\varepsilon_i$, we may also write:

$$\rho_{a0} = \frac{\varepsilon_a + j\sqrt{(\varepsilon_r + j\varepsilon_i)(\varepsilon_a - \varepsilon_b) - \varepsilon_a\varepsilon_b}}{\varepsilon_a - j\sqrt{(\varepsilon_r + j\varepsilon_i)(\varepsilon_a - \varepsilon_b) - \varepsilon_a\varepsilon_b}} \equiv -b_0 + ja_0 \tag{8.5.16}$$

which serves as the definition of b_0, a_0. We also write:

$$\rho_{a0}^{-1} = \frac{\varepsilon_a - j\sqrt{(\varepsilon_r + j\varepsilon_i)(\varepsilon_a - \varepsilon_b) - \varepsilon_a\varepsilon_b}}{\varepsilon_a + j\sqrt{(\varepsilon_r + j\varepsilon_i)(\varepsilon_a - \varepsilon_b) - \varepsilon_a\varepsilon_b}} = -\frac{b_0 + ja_0}{b_0^2 + a_0^2} \equiv -b_1 - ja_1 \tag{8.5.17}$$

We define also the wavenumber shifts that appear in the denominator and numerator of (8.5.15) as follows:

$$\begin{aligned} \bar{k}_{x0} &= -\rho_{a0}K_0 e^{-2jk_{z0}d} = (b_0 - ja_0)K_0 e^{-2jk_{z0}d} \equiv \bar{\beta}_{x0} - j\bar{\alpha}_{x0} \\ \bar{k}_{x1} &= -\rho_{a0}^{-1}K_0 e^{-2jk_{z0}d} = (b_1 + ja_1)K_0 e^{-2jk_{z0}d} \equiv \bar{\beta}_{x1} + j\bar{\alpha}_{x1} \end{aligned} \tag{8.5.18}$$

Then, Eq. (8.5.15) becomes, replacing $k_{x0} = \beta_{x0} - j\alpha_{x0}$

$$\Gamma = \rho_{a0}\,\frac{k_x - k_{x0} - \bar{k}_{x1}}{k_x - k_{x0} - \bar{k}_{x0}} = \rho_{a0}\,\frac{(k_x - \beta_{x0} - \bar{\beta}_{x1}) + j(\alpha_{x0} - \bar{\alpha}_{x1})}{(k_x - \beta_{x0} - \bar{\beta}_{x0}) + j(\alpha_{x0} + \bar{\alpha}_{x0})} \tag{8.5.19}$$

resulting in the reflectance:

$$|\Gamma|^2 = |\rho_{a0}|^2\,\frac{(k_x - \beta_{x0} - \bar{\beta}_{x1})^2 + (\alpha_{x0} - \bar{\alpha}_{x1})^2}{(k_x - \beta_{x0} - \bar{\beta}_{x0})^2 + (\alpha_{x0} + \bar{\alpha}_{x0})^2} \tag{8.5.20}$$

The shifted resonance wavenumber is determined from the denominator of (8.5.19), that is, $k_{x,\text{res}} = \beta_{x0} + \bar{\beta}_{x0}$. The resonance angle is determined by the matching condition:

$$\boxed{k_x = k_0 n_a \sin\theta_{\text{res}} = k_{x,\text{res}} = \beta_{x0} + \bar{\beta}_{x0}} \tag{8.5.21}$$

The minimum value of $|\Gamma|^2$ at resonance is obtained by setting $k_x = \beta_{x0} + \bar{\beta}_{x0}$:

$$|\Gamma|^2_{\text{min}} = |\rho_{a0}|^2\,\frac{(\bar{\beta}_{x0} - \bar{\beta}_{x1})^2 + (\alpha_{x0} - \bar{\alpha}_{x1})^2}{(\alpha_{x0} + \bar{\alpha}_{x0})^2} \tag{8.5.22}$$

We will see below that $\bar{\beta}_{x0}$ and $\bar{\beta}_{x1}$ are approximately equal, and so are $\bar{\alpha}_{x0}$ and $\bar{\alpha}_{x1}$. The optimum thickness for the metal layer is obtained by minimizing the numerator of $|\Gamma|^2_{\text{min}}$ by imposing the condition $\alpha_{x0} = \bar{\alpha}_{x1}$. This condition can be solved for d.

The angle width is obtained by solving for the left and right bandedge wavenumbers, say $k_{x,\pm}$, from the 3-dB condition:

$$|\Gamma|^2 = |\rho_{a0}|^2\,\frac{(k_x - \beta_{x0} - \bar{\beta}_{x1})^2 + (\alpha_{x0} - \bar{\alpha}_{x1})^2}{(k_x - \beta_{x0} - \bar{\beta}_{x0})^2 + (\alpha_{x0} + \bar{\alpha}_{x0})^2} = \frac{1}{2} \tag{8.5.23}$$

and then obtaining the left/right 3-dB angles by solving $k_0 n_a \sin\theta_\pm = k_{x,\pm}$.

Although Eqs. (8.5.16)–(8.5.23) can be easily implemented numerically, they are unnecessarily complicated. A further simplification can be made by replacing the quantities K_0, ρ_{a0}, and k_{z0} by their lossless values obtained by setting $\epsilon_i = 0$. This makes ρ_{a0} a unimodular complex number so that $\rho_{a0}^{-1} = \rho_{a0}^*$. We have then the approximations:

$$K_0 = k_0\left(\frac{2}{\epsilon_r + \epsilon_b}\right)\left(\frac{\epsilon_r\epsilon_b}{\epsilon_r - \epsilon_b}\right)^{3/2}$$

$$\rho_{a0} = \frac{\epsilon_a + j\sqrt{\epsilon_r(\epsilon_a - \epsilon_b) - \epsilon_a\epsilon_b}}{\epsilon_a - j\sqrt{\epsilon_r(\epsilon_a - \epsilon_b) - \epsilon_a\epsilon_b}} \equiv -b_0 + ja_0\,, \quad \rho_{a0}^{-1} = -b_0 - ja_0 \tag{8.5.24}$$

$$k_{z0} = -j\alpha_{z0}\,, \quad \alpha_{z0} = \frac{k_0\epsilon_r}{\sqrt{\epsilon_r - \epsilon_b}}$$

so that

$$b_0 = \frac{\epsilon_r(\epsilon_a - \epsilon_b) - \epsilon_a(\epsilon_a + \epsilon_b)}{(\epsilon_a - \epsilon_b)(\epsilon_r + \epsilon_b)}\,, \quad a_0 = \frac{2\epsilon_a\sqrt{\epsilon_r(\epsilon_a - \epsilon_b) - \epsilon_a\epsilon_b}}{(\epsilon_a - \epsilon_b)(\epsilon_r + \epsilon_b)} \tag{8.5.25}$$

The wavenumber shifts (8.5.18) then become:

$$\begin{aligned}\bar{k}_{x0} &= (b_0 - ja_0)K_0 e^{-2\alpha_{z0}d} = \bar{\beta}_{x0} - j\bar{\alpha}_{x0}\\ \bar{k}_{x1} &= (b_0 + ja_0)K_0 e^{-2\alpha_{z0}d} = \bar{\beta}_{x0} + j\bar{\alpha}_{x0} = \bar{k}_{x0}^*\end{aligned} \tag{8.5.26}$$

with

$$\bar{\beta}_{x0} = b_0 K_0 e^{-2\alpha_{z0}d}\,, \quad \bar{\alpha}_{x0} = a_0 K_0 e^{-2\alpha_{z0}d} \tag{8.5.27}$$

Then, the reflectance becomes in the neighborhood of the resonance:

$$|\Gamma|^2 = \frac{(k_x - \beta_{x0} - \bar{\beta}_{x0})^2 + (\alpha_{x0} - \bar{\alpha}_{x0})^2}{(k_x - \beta_{x0} - \bar{\beta}_{x0})^2 + (\alpha_{x0} + \bar{\alpha}_{x0})^2} \tag{8.5.28}$$

with a minimum value:

$$|\Gamma|^2_{\text{min}} = \frac{(\alpha_{x0} - \bar{\alpha}_{x0})^2}{(\alpha_{x0} + \bar{\alpha}_{x0})^2} \tag{8.5.29}$$

In this approximation, the resonance angle is determined from:

$$\boxed{k_0 n_a \sin\theta_{\text{res}} = k_{\text{res}} = \beta_{x0} + \bar{\beta}_{x0} = k_0\sqrt{\frac{\epsilon_r\epsilon_b}{\epsilon_r - \epsilon_b}} + b_0 K_0 e^{-2\alpha_{z0}d}} \tag{8.5.30}$$

Since the second term on the right-hand side represents a small correction, a necessary condition that such a resonance angle would exist is obtained by setting $\theta_{\text{res}} = 90^{\text{o}}$ and ignoring the second term:

$$n_a > \sqrt{\frac{\epsilon_r\epsilon_b}{\epsilon_r - \epsilon_b}} \equiv n_a^{\text{min}} \tag{8.5.31}$$

For example, for the parameters of Fig. 8.5.3, the minimum acceptable refractive index n_a would be $n_a^{\text{min}} = 1.033$. Thus, using a glass prism with $n_a = 1.5$ is more than adequate. If the right medium is water instead of air with $n_b = 1.33$, then $n_a^{\text{min}} = 1.41$, which comes close to the prism choice. The 3-dB angles are obtained by solving

$$|\Gamma|^2 = \frac{(k_x - k_{\text{res}})^2 + (\alpha_{x0} - \bar{\alpha}_{x0})^2}{(k_x - k_{\text{res}})^2 + (\alpha_{x0} + \bar{\alpha}_{x0})^2} = \frac{1}{2}$$

with solution $k_{x,\pm} = k_{\text{res}} \pm \sqrt{6\alpha_{x0}\bar{\alpha}_{x0} - \alpha_{x0}^2 - \bar{\alpha}_{x0}^2}$, or

$$k_0 n_a \sin\theta_\pm = k_0 n_a \sin\theta_{\text{res}} \pm \sqrt{6\alpha_{x0}\bar{\alpha}_{x0} - \alpha_{x0}^2 - \bar{\alpha}_{x0}^2} \tag{8.5.32}$$

The angle width shown on Fig. 8.5.3 was calculated by $\Delta\theta = \theta_+ - \theta_-$ using (8.5.32). The optimum thickness d_{opt} is obtained from the condition $\alpha_{x0} = \bar{\alpha}_{x0}$, which drives $|\Gamma|^2_{\text{min}}$ to zero. This condition requires that $\alpha_{x0} = a_0 K_0 e^{-2\alpha_{z0}d}$, with solution:

$$d_{\text{opt}} = \frac{1}{2\alpha_{z0}}\ln\left(\frac{a_0 K_0}{\alpha_{x0}}\right) = \frac{\lambda_0}{4\pi}\,\frac{\sqrt{\epsilon_r - \epsilon_b}}{\epsilon_r}\ln\left(\frac{4a_0\epsilon_r^2}{\epsilon_i(\epsilon_r + \epsilon_b)}\right) \tag{8.5.33}$$

where we replaced α_{x0} from Eq. (8.5.5). For the same parameters of Fig. 8.5.3, we calculate the optimum thickness to be $d_{\text{opt}} = 56.08$ nm, resulting in the new resonance angle of $\theta_{\text{res}} = 43.55^{\text{o}}$, and angle-width $\Delta\theta = 0.227^{\text{o}}$. Fig. 8.5.6 shows the reflectance in this case. The above approximations for the angle-width are not perfect, but they are adequate.

One of the current uses of surface plasmon resonance is the detection of the presence of chemical and biological agents. This application makes use of the fact that the

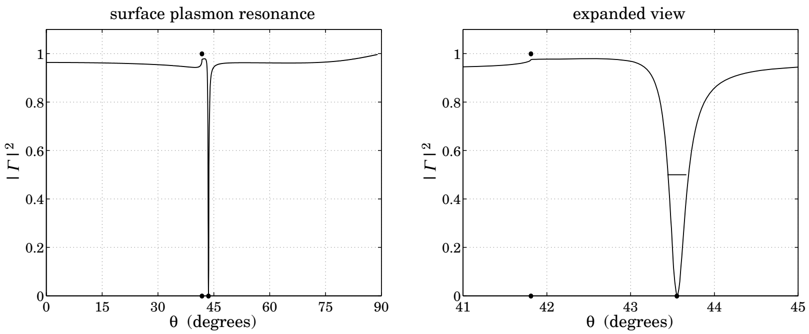


Fig. 8.5.6 Surface plasmon resonance at the optimum thickness $d = d_{\text{opt}}$.

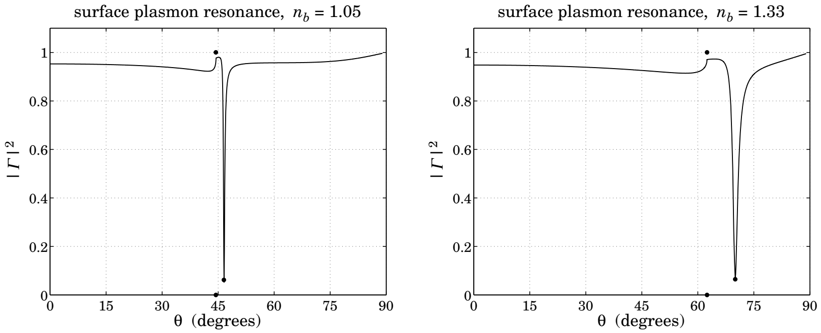


Fig. 8.5.7 Shift of the resonance angle with the refractive index n_b.

resonance angle θ_{res} is very sensitive to the dielectric constant of the medium n_b. For example, Fig. 8.5.7 shows the shift in the resonance angle for the two cases $n_b = 1.05$ and $n_b = 1.33$ (water). Using the same data as Fig. 8.5.3, the corresponding angles and widths were $\theta_{\text{res}} = 46.57^{\text{o}}$, $\Delta\theta = 0.349^{\text{o}}$ and $\theta_{\text{res}} = 70^{\text{o}}$, $\Delta\theta = 1.531^{\text{o}}$, respectively.

A number of applications of surface plasmons were mentioned in Sec. 7.7, such as nanophotonics and biosensors. The reader is referred to [553–591] for further reading.

8.6 Perfect Lens in Negative-Index Media

The perfect lens property of negative-index media was originally discussed by Veselago [353], who showed that a slab with $\epsilon = -\epsilon_0$ and $\mu = -\mu_0$, and hence with refractive index $n = -1$, can focus perfectly a point-source of light. More recently, Pendry [360] showed that such a slab can also amplify the evanescent waves from an object, and completely restore the object's spatial frequencies on the other side of the slab. The possibility of overcoming the diffraction limit and improving resolution with such a lens has generated a huge interest in the literature [353–434].

Fig. 8.6.1 shows the perfect lens property. Consider a ray emanating from an object at distance z_0 to the left of the slab ($z = -z_0$). Assuming vacuum on either side of the slab ($n_a = n_b = 1$), Snel's law, implies that the angle of incidence will be equal to the angle of refraction, bending in the same direction of the normal as the incident ray. Indeed, because $n_a = 1$ and $n = -1$, we have:

$$n_a \sin\theta_a = n\sin\theta \quad\Rightarrow\quad \sin\theta_a = -\sin\theta \quad\Rightarrow\quad \theta_a = -\theta$$

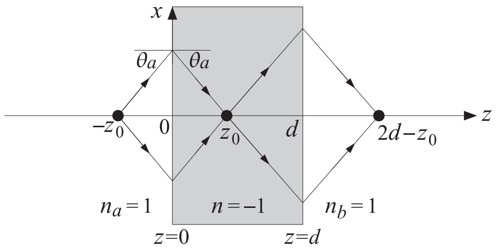


Fig. 8.6.1 Perfect lens property of a negative-index medium with $n = -1$

Moreover, $\eta = \sqrt{\mu/\epsilon} = \sqrt{\mu_0/\epsilon_0} = \eta_0$ and the slab is matched to the vacuum. Therefore, there will be no reflected ray at the left and the right interfaces. Indeed, the TE and TM reflection coefficients at the left interface vanish at any angle, for example, we have for the TM case, noting that $\cos\theta = \cos(-\theta_a) = \cos\theta_a$:

$$\rho_{TM} = \frac{\eta\cos\theta - \eta_0\cos\theta_a}{\eta\cos\theta + \eta_0\cos\theta_a} = \frac{\cos\theta - \cos\theta_a}{\cos\theta + \cos\theta_a} = 0$$

Assuming that $z_0 < d$, where d is the slab thickness, it can be seen from the geometry of Fig. 8.6.1 that the refracted rays will refocus at the point $z = z_0$ within the slab and then continue on to the right interface and refocus again at a distance $d - z_0$ from the slab, that is, at coordinate $z = 2d - z_0$.

Next, we examine the field solutions inside and outside the slab for propagating and for evanescent waves. For the TM case, the electric field will have the following form within the three regions of $z \le 0$, $0 \le z \le d$, and $z \ge d$:

$$\mathbf{E} = \begin{cases} \left[E_0\left(\hat{\mathbf{x}} - \dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z z} + E_0\Gamma\left(\hat{\mathbf{x}} + \dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z z}\right]e^{-jk_x x}, & \text{for}\quad z \le 0 \\[2ex] \left[A_+\left(\hat{\mathbf{x}} - \dfrac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{-jk'_z z} + A_-\left(\hat{\mathbf{x}} + \dfrac{k_x}{k'_z}\,\hat{\mathbf{z}}\right)e^{jk'_z z}\right]e^{-jk_x x}, & \text{for}\quad 0 \le z \le d \\[2ex] E_0 T\left(\hat{\mathbf{x}} - \dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z(z-d)}\,e^{-jk_x x}, & \text{for}\quad z \ge d \end{cases} \tag{8.6.1}$$

where Γ, T denote the overall transverse reflection and transmission coefficients, and A_+, A_-, the transverse fields on the right-side of the left interface (i.e., at $z = 0+$). The corresponding magnetic field is:

$$\boldsymbol{H}=\begin{cases}\hat{\mathbf{y}}\,E_0\left(\dfrac{\omega\epsilon}{k_z}\right)\big[e^{-jk_zz}-\Gamma e^{jk_zz}\big]e^{-jk_xx}\,, & \text{for}\quad z\le 0\\[2ex] \hat{\mathbf{y}}\left(\dfrac{\omega\epsilon'}{k'_z}\right)\big[A_+e^{-jk'_zz}-A_-e^{jk'_zz}\big]e^{-jk_xx}\,, & \text{for}\quad 0\le z\le d\\[2ex] \hat{\mathbf{y}}\,E_0T\left(\dfrac{\omega\epsilon}{k_z}\right)e^{-jk_z(z-d)}e^{-jk_xx}\,, & \text{for}\quad z\ge d\end{cases}\tag{8.6.2}$$

where k_x is preserved across the interfaces, and k_z, k'_z must satisfy:

$$k_x^2+k_z^2=\omega^2\mu_0\epsilon_0\,,\quad k_x^2+k_z'^2=\omega^2\mu\epsilon\tag{8.6.3}$$

Thus, $k_z=\pm\sqrt{\omega^2\mu_0\epsilon_0-k_x^2}$ and $k'_z=\pm\sqrt{\omega^2\mu\epsilon-k_x^2}$. The choice of square root signs is discussed below. To include evanescent waves, we will define k_z by means of the evanescent square root, setting $k_0=\omega\sqrt{\mu_0\epsilon_0}$:

$$k_z=\text{sqrte}\big(k_0^2-k_x^2\big)=\begin{cases}\sqrt{k_0^2-k_x^2}\,, & \text{if}\quad k_x^2\le k_0^2\\ -j\sqrt{k_x^2-k_0^2}\,, & \text{if}\quad k_x^2\ge k_0^2\end{cases}\tag{8.6.4}$$

We saw in Sec. 7.16 that for a single interface between a positive- and a negative-index medium, and for propagating waves, we must have $k_z>0$ and $k'_z<0$ in order for the power transmitted into the negative-index medium to flow away from the interface. But in the case of a slab within which one could have both forward and backward waves, the choice of the sign of k'_z is not immediately obvious. In fact, it turns out that the field solution remains *invariant* under the substitution $k'_z\to -k'_z$, and therefore, one could choose either sign for k'_z. In particular, we could select it to be given also by its evanescent square root, where $n^2=\epsilon\mu/\epsilon_0\mu_0$:

$$k'_z=\text{sqrte}\big(k_0^2n^2-k_x^2\big)=\begin{cases}\sqrt{k_0^2n^2-k_x^2}\,, & \text{if}\quad k_x^2\le k_0^2n^2\\ -j\sqrt{k_x^2-k_0^2n^2}\,, & \text{if}\quad k_x^2\ge k_0^2n^2\end{cases}\tag{8.6.5}$$

By matching the boundary conditions at the two interfaces $z=0$ and $z=d$, the parameters $\Gamma, A_\pm, T$ are obtained from the usual transfer matrices (see Sec. 8.1):

$$\begin{aligned}\begin{bmatrix}E_0\\ E_0\Gamma\end{bmatrix}&=\frac{1}{1+\rho_{TM}}\begin{bmatrix}1 & \rho_{TM}\\ \rho_{TM} & 1\end{bmatrix}\begin{bmatrix}A_+\\ A_-\end{bmatrix}\\ \begin{bmatrix}A_+\\ A_-\end{bmatrix}&=\begin{bmatrix}e^{jk'_zd} & 0\\ 0 & e^{-jk'_zd}\end{bmatrix}\frac{1}{1-\rho_{TM}}\begin{bmatrix}1 & -\rho_{TM}\\ -\rho_{TM} & 1\end{bmatrix}\begin{bmatrix}E_0T\\ 0\end{bmatrix}\end{aligned}\tag{8.6.6}$$

where,

$$\rho_{TM}=\frac{k'_z\epsilon-k_z\epsilon'}{k'_z\epsilon+k_z\epsilon'}=\frac{\zeta_{TM}-1}{\zeta_{TM}+1}\,,\quad \zeta_{TM}=\frac{\eta'_{TM}}{\eta_{TM}}=\frac{k'_z\epsilon}{k_z\epsilon'}\tag{8.6.7}$$

where ζ_{TM} is a normalized characteristic impedance. The solution of Eqs. (8.6.6) is then,

$$\begin{aligned}\Gamma&=\frac{\rho_{TM}\big(1-e^{-2jk'_zd}\big)}{1-\rho_{TM}^2e^{-2jk'_zd}}=\frac{(\zeta_{TM}^2-1)\,(1-e^{-2jk'_zd})}{(\zeta_{TM}+1)^2-(\zeta_{TM}-1)^2e^{-2jk'_zd}}\\ T&=\frac{(1-\rho_{TM}^2)\,e^{-jk'_zd}}{1-\rho_{TM}^2e^{-2jk'_zd}}=\frac{4\zeta_{TM}}{(\zeta_{TM}+1)^2e^{jk'_zd}-(\zeta_{TM}-1)^2e^{-jk'_zd}}\end{aligned}\tag{8.6.8}$$

Similarly, the coefficients $A_\pm$ are found to be:

$$\begin{aligned}A_+&=\frac{1-\rho_{TM}\Gamma}{1-\rho_{TM}}=\frac{1}{2}\big[1+\zeta_{TM}+(1-\zeta_{TM})\Gamma\big]E_0\\ A_-&=\frac{-\rho_{TM}+\Gamma}{1-\rho_{TM}}=\frac{1}{2}\big[1-\zeta_{TM}+(1+\zeta_{TM})\Gamma\big]E_0\end{aligned}\tag{8.6.9}$$

The TE case is obtained from the TM case by a duality transformation, that is, by the replacements, $\boldsymbol{E}\to\boldsymbol{H}$, $\boldsymbol{H}\to-\boldsymbol{E}$, $\epsilon\to\mu$, $\epsilon'\to\mu'$, and $\rho_{TM}\to\rho_{TE}$, where

$$\rho_{TE}=\frac{k_z\mu'-k'_z\mu}{k_z\mu'+k'_z\mu}=\frac{\zeta_{TE}-1}{\zeta_{TE}+1}\,,\quad \zeta_{TE}=\frac{\eta'_{TE}}{\eta_{TE}}=\frac{k_z\mu'}{k'_z\mu}$$

The invariance under the transformation $k'_z\to-k'_z$ follows from these solutions. For example, noting that $\zeta_{TM}\to-\zeta_{TM}$ under this transformation, we have:

$$\Gamma(-k'_z)=\frac{(\zeta_{TM}^2-1)\,(1-e^{2jk'_zd})}{(-\zeta_{TM}+1)^2-(-\zeta_{TM}-1)^2e^{2jk'_zd}}=\frac{(\zeta_{TM}^2-1)\,(1-e^{-2jk'_zd})}{(\zeta_{TM}+1)^2-(\zeta_{TM}-1)^2e^{-2jk'_zd}}=\Gamma(k'_z)$$

Similarly, we find $T(-k'_z)=T(k'_z)$ and $A_\pm(-k'_z)=A_\mp(k'_z)$. These imply that the field solutions remain invariant. For example, the electric field inside the slab will be:

$$\begin{aligned}\boldsymbol{E}(z,-k'_z)&=\Big[A_+(-k'_z)\Big(\hat{\mathbf{x}}-\frac{k_x}{-k'_z}\,\hat{\mathbf{z}}\Big)e^{-j(-k'_z)z}+A_-(-k'_z)\Big(\hat{\mathbf{x}}+\frac{k_x}{-k'_z}\,\hat{\mathbf{z}}\Big)e^{j(-k'_z)z}\Big]e^{-jk_xx}\\ &=\Big[A_+(k'_z)\Big(\hat{\mathbf{x}}-\frac{k_x}{k'_z}\,\hat{\mathbf{z}}\Big)e^{-jk'_zz}+A_-(k'_z)\Big(\hat{\mathbf{x}}+\frac{k_x}{k'_z}\,\hat{\mathbf{z}}\Big)e^{jk'_zz}\Big]e^{-jk_xx}=\boldsymbol{E}(z,+k'_z)\end{aligned}$$

Similarly, we have for the magnetic field inside the slab:

$$\begin{aligned}\boldsymbol{H}(z,-k'_z)&=\hat{\mathbf{y}}\left(\frac{\omega\epsilon'}{-k'_z}\right)\big[A_+(-k'_z)e^{-j(-k'_z)z}-A_-(-k'_z)e^{j(-k'_z)z}\big]e^{-jk_xx}\\ &=\hat{\mathbf{y}}\left(\frac{\omega\epsilon'}{k'_z}\right)\big[A_+(k'_z)e^{-jk'_zz}-A_-(k'_z)e^{jk'_zz}\big]e^{-jk_xx}=\boldsymbol{H}(z,+k'_z)\end{aligned}$$

Next, we apply these results to the case $\mu=-\mu_0$ and $\epsilon=-\epsilon_0$, having $n=-1$. It follows from Eq. (8.6.5) that $k'_z=\mp k_z$ with k_z given by (8.6.4). In this case, $\zeta_{TM}=k'_z\epsilon/k_z\epsilon'=-k'_z/k_z=\pm1$. Then, Eq. (8.6.8) implies that $\Gamma=0$ for either choice of sign. Similarly, we have $T=e^{jk_zd}$, again for either sign of ζ_{TM}:

$$T=e^{jk_zd}=\begin{cases}e^{jk_zd}\,, & \text{if}\quad k_x^2\le k_0^2,\quad k_z=\sqrt{k_0^2-k_x^2}\\ e^{\alpha_zd}\,, & \text{if}\quad k_x^2\ge k_0^2,\quad k_z=-j\sqrt{k_x^2-k_0^2}\equiv-j\alpha_z\end{cases}\tag{8.6.10}$$

Thus, the negative-index medium amplifies the transmitted evanescent waves, which was Pendry's observation [360]. The two choices for k'_z lead to the $A_\pm$ coefficients:

$$\begin{aligned}k'_z&=-k_z\quad\Rightarrow\quad\zeta_{TM}=+1\quad\Rightarrow\quad A_+=E_0\,,\quad A_-=0\\ k'_z&=+k_z\quad\Rightarrow\quad\zeta_{TM}=-1\quad\Rightarrow\quad A_+=0\,,\quad A_-=E_0\end{aligned}\tag{8.6.11}$$

For either choice, the field solutions are the same. Indeed, inserting either set of A_+, A_- into Eqs. (8.6.1) and (8.6.2), and using (8.6.10), we find:

$$\boldsymbol{E}=\begin{cases} E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{-jk_zz}e^{-jk_xx}\,, & \text{for}\quad z\le 0\\[2ex] E_0\Big(\hat{\mathbf{x}}+\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{jk_zz}e^{-jk_xx}\,, & \text{for}\quad 0\le z\le d\\[2ex] E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{-jk_z(z-2d)}e^{-jk_xx}\,, & \text{for}\quad z\ge d \end{cases} \tag{8.6.12}$$

and the corresponding magnetic field:

$$\boldsymbol{H}=\begin{cases} \hat{\mathbf{y}}\,E_0\left(\dfrac{\omega\epsilon_0}{k_z}\right)e^{-jk_zz}e^{-jk_xx}\,, & \text{for}\quad z\le 0\\[2ex] \hat{\mathbf{y}}\,E_0\left(\dfrac{\omega\epsilon_0}{k_z}\right)e^{jk_zz}e^{-jk_xx}\,, & \text{for}\quad 0\le z\le d\\[2ex] \hat{\mathbf{y}}\,E_0\left(\dfrac{\omega\epsilon_0}{k_z}\right)e^{-jk_z(z-2d)}e^{-jk_xx}\,, & \text{for}\quad z\ge d \end{cases} \tag{8.6.13}$$

The solution effectively corresponds to the choice $k_z' = -k_z$ and is valid for both propagating and evanescent waves with k_z given by (8.6.4). In Eq. (8.6.12) the constant E_0 refers to the value of the transverse electric field at $z = 0$. Changing the reference point to $z = -z_0$ at the left of the slab as shown in Fig. 8.6.1, amounts to replacing $E_0 \to E_0 e^{-jk_zz_0}$. Then, (8.6.12) reads:

$$\boldsymbol{E}=\begin{cases} E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{-jk_z(z+z_0)}e^{-jk_xx}\,, & \text{for}\quad -z_0\le z\le 0\\[2ex] E_0\Big(\hat{\mathbf{x}}+\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{jk_z(z-z_0)}e^{-jk_xx}\,, & \text{for}\quad 0\le z\le d\\[2ex] E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{k_z}\,\hat{\mathbf{z}}\Big)e^{-jk_z(z-2d+z_0)}e^{-jk_xx}\,, & \text{for}\quad z\ge d \end{cases} \tag{8.6.14}$$

Setting $k_z = -\alpha_z$ as in (8.6.4), we find the for the evanescent fields:

$$\boldsymbol{E}=\begin{cases} E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{-j\alpha_z}\,\hat{\mathbf{z}}\Big)e^{-\alpha_z(z+z_0)}e^{-jk_xx}\,, & \text{for}\quad -z_0\le z\le 0\\[2ex] E_0\Big(\hat{\mathbf{x}}+\dfrac{k_x}{-j\alpha_z}\,\hat{\mathbf{z}}\Big)e^{\alpha_z(z-z_0)}e^{-jk_xx}\,, & \text{for}\quad 0\le z\le d\\[2ex] E_0\Big(\hat{\mathbf{x}}-\dfrac{k_x}{-j\alpha_z}\,\hat{\mathbf{z}}\Big)e^{-\alpha_z(z-2d+z_0)}e^{-jk_xx}\,, & \text{for}\quad z\ge d \end{cases} \tag{8.6.15}$$

The field is amplified inside the slab. The propagation factors along the z-direction agree at the points $z = -z_0$, $z = z_0$, and $z = 2d - z_0$,

$$\begin{aligned} e^{-jk_z(z+z_0)}\Big|_{z=-z_0} &= e^{jk_z(z-z_0)}\Big|_{z=z_0} = e^{-jk_z(z-2d+z_0)}\Big|_{z=2d-z_0} = 1\\ e^{-\alpha_z(z+z_0)}\Big|_{z=-z_0} &= e^{\alpha_z(z-z_0)}\Big|_{z=z_0} = e^{-\alpha_z(z-2d+z_0)}\Big|_{z=2d-z_0} = 1 \end{aligned} \tag{8.6.16}$$

which imply the complete restoration of the source at the focal points inside and to the right of the slab:

$$\boldsymbol{E}(x,z)\,\big|_{z=-z_0} = \boldsymbol{E}(x,z)\,\big|_{z=z_0} = \boldsymbol{E}(x,z)\,\big|_{z=2d-z_0} \tag{8.6.17}$$

Fig. 8.6.2 shows a plot of the evanescent component $E_x(z)$ of Eq. (8.6.15) versus distance z inside and outside the slab.

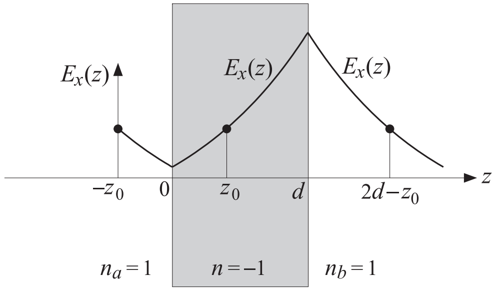


Fig. 8.6.2 Evanesenct wave amplification inside a negative-index medium.

Using the plane-wave spectrum representation of Sec. 17.17, a more general (single-frequency) solution can be built by superposition of the plane waves (8.6.14) and (8.6.15). If the field at the image plane $z = -z_0$ has the general representation:

$$\boldsymbol{E}(x,-z_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}E_0(k_x)\left(\hat{\mathbf{x}}-\frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_xx}\,dk_x \tag{8.6.18}$$

where the integral over k_x includes both propagating and evanescent modes and k_z is given by (8.6.4), then, then field in the three regions to the left of, inside, and to the right of the slab will have the form:

$$\begin{aligned} \boldsymbol{E}(x,z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}E_0(k_x)\left(\hat{\mathbf{x}}-\frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z(z+z_0)}e^{-jk_xx}\,dk_x\,, && \text{for}\quad -z_0\le z\le 0\\ \boldsymbol{E}(x,z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}E_0(k_x)\left(\hat{\mathbf{x}}+\frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{jk_z(z-z_0)}e^{-jk_xx}\,dk_x\,, && \text{for}\quad 0\le z\le d\\ \boldsymbol{E}(x,z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}E_0(k_x)\left(\hat{\mathbf{x}}-\frac{k_x}{k_z}\,\hat{\mathbf{z}}\right)e^{-jk_z(z-2d+z_0)}e^{-jk_xx}\,dk_x\,, && \text{for}\quad z\ge d \end{aligned}$$

It is evident that Eq. (8.6.17) is still satisfied, showing the perfect reconstruction of the object field at the two image planes.

The perfect lens property is highly sensitive to the deviations from the ideal values of $\epsilon = -\epsilon_0$ and $\mu = -\mu_0$, and to the presence of losses. Fig. 8.6.3 plots the transmittance in dB, that is, the quantity $10\log_{10}|Te^{-jk_zd}|^2$ versus k_x, with T computed from Eq. (8.6.8) for different values of ϵ,μ and for $d = 0.2\lambda = 0.2(2\pi/k_0)$. In the ideal case, because of the result (8.6.10), we have $|Te^{-jk_zd}| = 1$ for both propagating and evanescent values of k_x, that is, the transmittance is flat (at 0 dB) for all k_x.

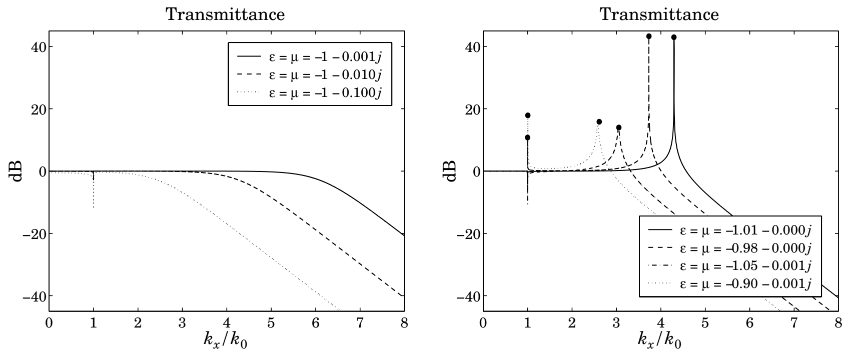


Fig. 8.6.3 Transmittance under non-ideal conditions (ϵ, μ are in units of ϵ_0, μ_0).

The left graph shows the effect of losses while keeping the real parts of ϵ, μ at the ideal values $-\epsilon_0, -\mu_0$. In the presence of losses, the transmittance acts like a lowpass filter in the spatial frequency k_x.

The right graph shows the effect of the deviation of the real parts of ϵ, μ from the ideal values. If the real parts deviate, even slightly, from $-\epsilon_0, -\mu_0$, the transmittance develops resonance peaks, which are related to the excitation of surface plasmons at the two interfaces of the slab [369,370]. The peaks are due to the poles of the denominator of T in Eq. (8.6.8), that is, the roots of

$$1 - \rho_{TM}^2 e^{-2jk_z'd} = 0 \quad\Rightarrow\quad e^{2jk_z'd} = \rho_{TM}^2 \quad\Rightarrow\quad e^{jk_z'd} = \pm\rho_{TM}$$

For evanescent k_x, we may replace $k_z = -j\alpha_z$ and $k_z' = -j\alpha_z'$, where $\alpha_z = \sqrt{k_x^2 - k_0^2}$ and $\alpha_z' = \sqrt{k_x^2 - k_0^2 n^2}$, and obtain the conditions:

$$e^{\alpha_z' d} = \pm\rho_{TM} = \pm\frac{\alpha_z'\epsilon_0 - \alpha_z\epsilon}{\alpha_z'\epsilon_0 + \alpha_z\epsilon} \tag{8.6.19}$$

These are equivalent to [369,370]:

$$\tanh\left(\frac{\alpha_z' d}{2}\right) = -\frac{\alpha_z\epsilon}{\alpha_z'\epsilon_0}\,, \quad \tanh\left(\frac{\alpha_z' d}{2}\right) = -\frac{\alpha_z'\epsilon_0}{\alpha_z\epsilon} \tag{8.6.20}$$

For $k_x \gg k_0$, we may replace $\alpha_z = \alpha_z' \simeq k_x$ in (8.6.19) in order to get en estimate of the resonant k_x:

$$e^{k_{x,\mathrm{res}}d} = \pm\frac{\epsilon_0 - \epsilon}{\epsilon_0 + \epsilon} \quad\Rightarrow\quad e^{\mathrm{Re}(k_{x,\mathrm{res}})d} = \left|\frac{\epsilon_0 - \epsilon}{\epsilon_0 + \epsilon}\right| \quad\Rightarrow\quad \mathrm{Re}(k_{x,\mathrm{res}}) = \frac{1}{d}\left|\frac{\epsilon_0 - \epsilon}{\epsilon_0 + \epsilon}\right| \tag{8.6.21}$$

and for the TE case, we must replace ϵs by μs. The value $k_x = \mathrm{Re}(k_{x,\mathrm{res}})$ represents the highest achievable resolution by the slab, with the smallest resolvable transverse distance being of the order of $\Delta x = 1/\mathrm{Re}(k_{x,\mathrm{res}})$.

If ϵ is real-valued and near $-\epsilon_0$, then, $k_{x,\mathrm{res}}$ is real and there will be an infinite resonance peak at $k_x = k_{x,\mathrm{res}}$. This seen in the above figure in the first two cases of $\epsilon/\epsilon_0 = \mu/\mu_0 = -1.01$ and $\epsilon/\epsilon_0 = \mu/\mu_0 = -0.98$ (the apparent finite height of these two peaks is due to the finite grid of k_x values in the graph.)

The last two cases have complex-valued ϵ, μ with a small imaginary part, with the resulting peaks being finite. In all cases, the peak locations $k_x = \mathrm{Re}(k_{x,\mathrm{res}})$—obtained by solving Eqs. (8.6.20) numerically for $k_{x,\mathrm{res}}$—are indicated in the graphs by bullets placed at the peak maxima. The numerical solutions were obtained by the following iterative procedures, initialized at the approximate (complex-valued) solution of (8.6.21):

```
initialize:  $k_x = \frac{1}{d}\ln\left(\frac{\epsilon_0-\epsilon}{\epsilon_0+\epsilon}\right)$
for $i = 1, 2, \dots, N_{\mathrm{iter}}$, do:
    $\alpha_z' = \sqrt{k_x^2 - k_0^2 n^2}$
    $\alpha_z = -\frac{\epsilon_0}{\epsilon}\,\alpha_z'\tanh\left(\frac{\alpha_z' d}{2}\right)$
    $k_x = \sqrt{\alpha_z^2 + k_0^2}$
```

or,

```
$k_x = \frac{1}{d}\ln\left(-\frac{\epsilon_0-\epsilon}{\epsilon_0+\epsilon}\right)$ ,   $\alpha_z' = \sqrt{k_x^2 - k_0^2 n^2}$
for $i = 1, 2, \dots, N_{\mathrm{iter}}$, do:
    $\alpha_z = \sqrt{k_x^2 - k_0^2}$
    $\alpha_z' = -\frac{\epsilon}{\epsilon_0}\,\alpha_z\tanh\left(\frac{\alpha_z' d}{2}\right)$
    $k_x = \sqrt{\alpha_z'^2 + k_0^2 n^2}$
```

The number of iterations was typically $N_{\mathrm{iter}} = 30$. Both graphs of Fig. 8.6.3 also show dips at $k_x = k_0$. These are due to the zeros of the transmittance T arising from the numerator factor $(1 - \rho_{TM}^2)$ in (8.6.10). At $k_x = k_0$, we have $\alpha_z = 0$ and $\rho_{TM} = 1$, causing a zero in T. In addition to the zero at $k_x = k_0$, it is possible to also have poles in the vicinity of k_0, as indicated by the peaks and bullets in the graph. Fig. 8.6.4 shows an expanded view of the structure of T near k_0, with the k_x restricted in the narrow interval: $0.99k_0 \le k_x \le 1.01k_0$.

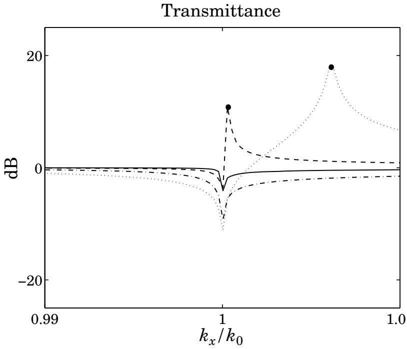


Fig. 8.6.4 Expanded view of the zero/pole behavior in the vicinity of $k_x = k_0$.

For last two cases depicted on this graph that have $|n^2| = |\epsilon\mu|/\epsilon_0\mu_0 \lesssim 1$, an approximate calculation of the pole locations near k_0 is as follows. Since $\alpha_z = \sqrt{k_x^2 - k_0^2}$ is small, and $\alpha_z' = \sqrt{\alpha_z^2 + k_0(1 - n^2)}$, we have to first order in α_z, $\alpha_z' \simeq k_0^2\sqrt{1 - n^2} \equiv \alpha_{z0}'$, which is itself small. Then, we apply Eq. (8.6.21) to get α_z and from it, the resonant $k_{x,\mathrm{res}}$:

$$\alpha_z = -\frac{\epsilon_0}{\epsilon}\,\alpha'_{z0}\tanh\left(\frac{\alpha'_{z0}d}{2}\right) \quad\Rightarrow\quad k_{x,\mathrm{res}} = \sqrt{\alpha_z^2 + k_0^2}$$

8.7 Antireflection Coatings at Oblique Incidence

Antireflection coatings are typically designed for normal incidence and then used over a limited range of oblique incidence, such as up to about 30°. As the angle of incidence increases, the antireflection band shifts towards lower wavelengths or higher frequencies. Any designed reflection zeros at normal incidence are no longer zeros at oblique incidence.

If a particular angle of incidence is preferred, it is possible to design the antireflection coating to match that angle. However, like the case of normal design, the effectiveness of this method will be over an angular width of approximately 30° about the preferred angle.

To appreciate the effects of oblique incidence, we look at the angular behavior of our normal-incidence designs presented in Figs. 6.2.1 and 6.2.3.

The first example was a two-layer design with refractive indices $n_a = 1$ (air), $n_1 = 1.38$ (magnesium fluoride), $n_2 = 2.45$ (bismuth oxide), and $n_b = 1.5$ (glass). The designed normalized optical lengths of the layers were $L_1 = 0.3294$ and $L_2 = 0.0453$ at $\lambda_0 = 550$ nm.

Fig. 8.7.1 shows the TE and TM reflectances $|\Gamma_{T1}(\lambda)|^2$ as functions of λ, for the incidence angles $\theta = 0^\circ, 20^\circ, 30^\circ, 40^\circ$.

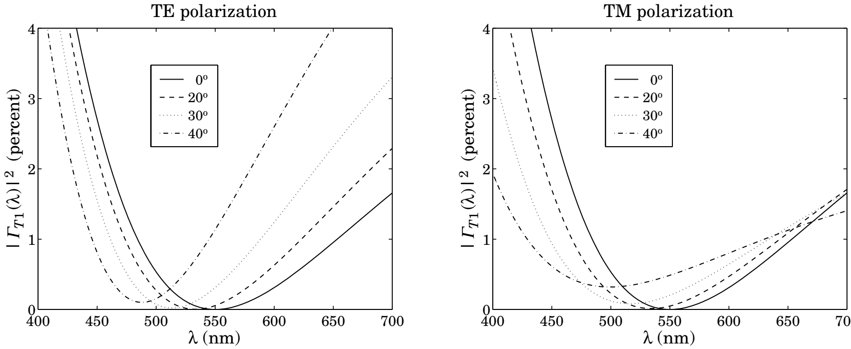


Fig. 8.7.1 Two-layer antireflection coating at oblique incidence.

We note the shifting of the responses towards lower wavelengths. The responses are fairly acceptable up to about 20°–30°. The typical MATLAB code used to generate these graphs was:

```
n = [1, 1.38, 2.45, 1.5]; L = [0.3294, 0.0453];
la0 = 550; la = linspace(400,700,101); pol='te';
```

```
G0  = abs(multidiel(n, L, la/la0)).^2 * 100;
G20 = abs(multidiel(n, L, la/la0, 20, pol)).^2 * 100;
G30 = abs(multidiel(n, L, la/la0, 30, pol)).^2 * 100;
G40 = abs(multidiel(n, L, la/la0, 40, pol)).^2 * 100;

plot(la, [G0; G20; G30; G40]);
```

As we mentioned above, the design can be matched at a particular angle of incidence. As an example, we choose $\theta_a = 30^\circ$ and redesign the two-layer structure.

The design equations are still (6.2.2) and (6.2.1), but with the replacement of n_i, ρ_i by their transverse values n_{Ti}, ρ_{Ti}, and the replacement of $k_1 l_1$, $k_2 l_2$ by the phase thicknesses at $\lambda = \lambda_0$, that is, $\delta_1 = 2\pi L_1\cos\theta_1$ and $\delta_2 = 2\pi L_2\cos\theta_2$. Moreover, we must choose to match the design either for TE or TM polarization.

Fig. 8.7.2 illustrates such a design. The upper left graph shows the TE reflectance matched at 30°. The designed optical thicknesses are in this case, $L_1 = 0.3509$ and $L_2 = 0.0528$. The upper right graph shows the corresponding TM reflectance, which cannot be matched simultaneously with the TE case.

The lower graphs show the same design, but now the TM reflectance is matched at 30°. The designed lengths were $L_1 = 0.3554$ and $L_2 = 0.0386$.

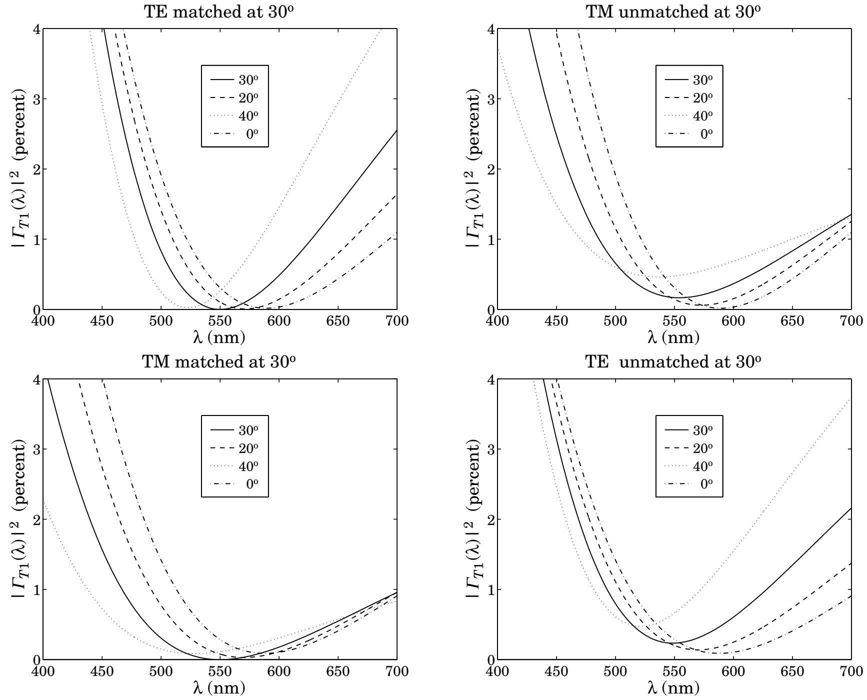


Fig. 8.7.2 Two-layer antireflection coating matched at 30 degrees.

The design steps are as follows. First, we calculate the refraction angles in all media from Eq. (8.1.1), $\theta_i = \text{asin}(n_a \sin\theta_a / n_i)$, for $i = a, 1, 2, b$. Then, assuming TE polarization, we calculate the TE refractive indices for all media $n_{Ti} = n_i \cos\theta_i$, $i = a, 1, 2, b$.

Then, we calculate the transverse reflection coefficients ρ_{Ti} from Eq. (8.1.3) and use them to solve Eq. (6.2.2) and (6.2.1) for the phase thicknesses δ_1, δ_2. Finally, we calculate the normalized optical lengths from $L_i = \delta_i / (2\pi \cos\theta_i)$, $i = 1, 2$. The following MATLAB code illustrates these steps:

```
n = [1, 1.38, 2.45, 1.5];
tha = 30;  thi = asin(na*sin(pi*tha/180)./n);

nt = n.*cos(thi);                    % for TM use nt = n./cos(thi)
r = n2r(nt);

c = sqrt((r(1)^2*(1-r(2)*r(3))^2 - (r(2)-r(3))^2)/(4*r(2)*r(3)*(1-r(1)^2)));
de2 = acos(c);
G2 = (r(2)+r(3)*exp(-2*j*de2))/(1 + r(2)*r(3)*exp(-2*j*de2));
de1 = (angle(G2) - pi - angle(r(1)))/2;
if de1 <0, de1 = de1 + 2*pi; end

L = [de1,de2]/2/pi;
L = L./cos(thi(2:3));

la0 = 550; la = linspace(400,700,401);

G30 = abs(multidiel(n, L, la/la0, 30, ’te’)).^2 * 100;
G20 = abs(multidiel(n, L, la/la0, 20, ’te’)).^2 * 100;
G40 = abs(multidiel(n, L, la/la0, 40, ’te’)).^2 * 100;
G0  = abs(multidiel(n, L, la/la0)).^2 * 100;

plot(la, [G30; G20; G40; G0]);
```

Our second example in Fig. 6.2.3 was a quarter-half-quarter 3-layer design with refractive indices $n_1 = 1$ (air), $n_1 = 1.38$ (magnesium fluoride), $n_2 = 2.2$ (zirconium oxide), $n_3 = 1.63$ (cerium fluoride), and $n_b = 1.5$ (glass). The optical lengths of the layers were $L_1 = L_3 = 0.25$ and $L_2 = 0.5$.

Fig. 8.7.3 shows the TE and TM reflectances $|\Gamma_{T1}(\lambda)|^2$ as functions of λ, for the incidence angles $\theta = 0^o, 20^o, 30^o, 40^o$.

The responses are fairly acceptable up to about 20^o–30^o, but are shifted towards lower wavelengths. The typical MATLAB code used to generate these graphs was:

```
n = [1, 1.38, 2.2, 1.63, 1.5]; L = [0.25, 0.50, 0.25];

la0 = 550; la = linspace(400,700,401);

G0  = abs(multidiel(n, L, la/la0)).^2 * 100;
G20 = abs(multidiel(n, L, la/la0, 20, ’te’)).^2 * 100;
G30 = abs(multidiel(n, L, la/la0, 30, ’te’)).^2 * 100;
G40 = abs(multidiel(n, L, la/la0, 40, ’te’)).^2 * 100;

plot(la, [G0; G20; G30; G40]);
```

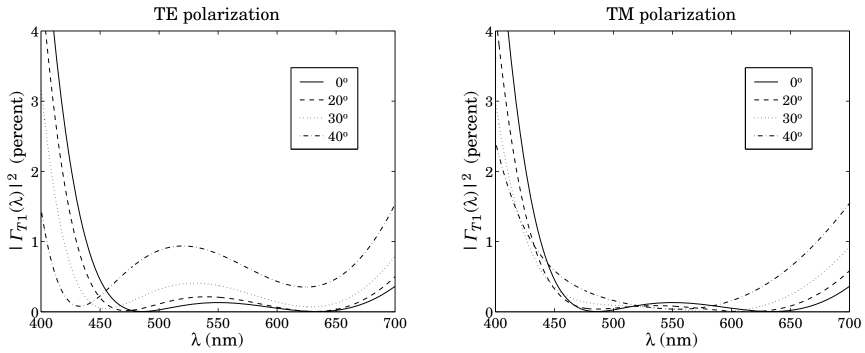


Fig. 8.7.3 Three-layer antireflection coating at oblique incidence.

8.8 Omnidirectional Dielectric Mirrors

Until recently, it was generally thought that it was impossible to have an omnidirectional dielectric mirror, that is, a mirror that is perfectly reflecting at all angles of incidence and for both TE and TM polarizations. However, such mirrors are possible and have recently been manufactured [736,737] and the conditions for their existence clarified [736–740].

We consider the same dielectric mirror structure of Sec. 6.3, consisting of alternating layers of high and low index. Fig. 8.8.1 shows such a structure under oblique incidence. There are N bilayers and a total of $M = 2N + 1$ single layers, starting and ending with a high-index layer.

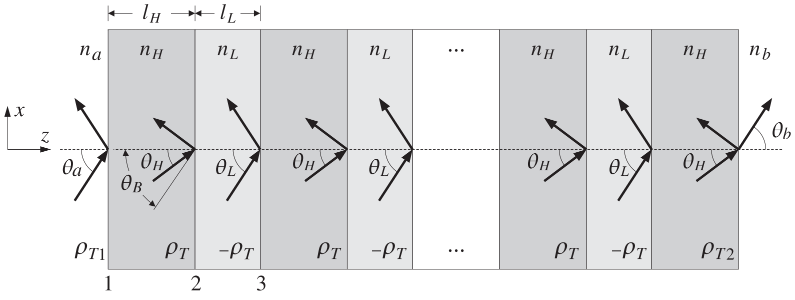


Fig. 8.8.1 Dielectric mirror at oblique incidence.

The incidence angles on each interface are related by Snel’s law:

$$n_a \sin\theta_a = n_H \sin\theta_H = n_L \sin\theta_L = n_b \sin\theta_b \tag{8.8.1}$$

The phase thicknesses within the high- and low-index layers are in normalized form:

$$\boxed{\delta_H = 2\pi\frac{f}{f_0}L_H\cos\theta_H\,,\quad \delta_L = 2\pi\frac{f}{f_0}L_L\cos\theta_L} \tag{8.8.2}$$

where $L_H = n_H l_H/\lambda_0$, $L_L = n_L l_L/\lambda_0$ are the optical thicknesses normalized to some λ_0, and $f_0 = c_0/\lambda_0$. Note also, $\cos\theta_i = \sqrt{1 - n_a^2\sin^2\theta_a/n_i^2}$, $i = H, L$.

A necessary (but not sufficient) condition for omnidirectional reflectivity for both polarizations is that the maximum angle of refraction $\theta_{H,\max}$ inside the first layer be less than the Brewster angle θ_B of the second interface, that is, the high-low interface, so that the Brewster angle can never be accessed by a wave incident on the first interface. If this condition is not satisfied, a TM wave would not be reflected at the second and all subsequent interfaces and will transmit through the structure.

Because $\sin\theta_{H,\max} = n_a/n_H$ and $\tan\theta_B = n_L/n_H$, or, $\sin\theta_B = n_L/\sqrt{n_H^2 + n_L^2}$, the condition $\theta_{H,\max} < \theta_B$, or the equivalent condition $\sin\theta_{H,\max} < \sin\theta_B$, can be written as $n_a/n_H < n_L/\sqrt{n_H^2 + n_L^2}$, or

$$\boxed{n_a < \frac{n_H n_L}{\sqrt{n_H^2 + n_L^2}}} \tag{8.8.3}$$

We note that the exact opposite of this condition is required in the design of multilayer Brewster polarizing beam splitters, discussed in the next section.

In addition to condition (8.8.3), in order to achieve omnidirectional reflectivity we must require that the high-reflectance bands have a *common* overlapping region for all incidence angles and for both polarizations.

To determine these bands, we note that the entire discussion of Sec. 6.3 carries through unchanged, provided we use the transverse reflection coefficients and transverse refractive indices. For example, the transverse version of the bilayer transition matrix of Eq. (6.3.5) will be:

$$F_T = \frac{1}{1-\rho_T^2}\begin{bmatrix} e^{j(\delta_H+\delta_L)} - \rho_T^2 e^{j(\delta_H-\delta_L)} & -2j\rho_T e^{-j\delta_H}\sin\delta_L \\ 2j\rho_T e^{j\delta_H}\sin\delta_L & e^{-j(\delta_H+\delta_L)} - \rho_T^2 e^{-j(\delta_H-\delta_L)} \end{bmatrix} \tag{8.8.4}$$

where $\rho_T = (n_{HT} - n_{LT})/(n_{HT} + n_{LT})$ and:

$$n_{HT} = \begin{cases} \dfrac{n_H}{\cos\theta_H} \\ n_H\cos\theta_H \end{cases} \qquad n_{LT} = \begin{cases} \dfrac{n_L}{\cos\theta_L} & \text{(TM polarization)} \\ n_L\cos\theta_L & \text{(TE polarization)} \end{cases} \tag{8.8.5}$$

Explicitly, we have for the two polarizations:

$$\rho_{TM} = \frac{n_H\cos\theta_L - n_L\cos\theta_H}{n_H\cos\theta_L + n_L\cos\theta_H}\,,\quad \rho_{TE} = \frac{n_H\cos\theta_H - n_L\cos\theta_L}{n_H\cos\theta_H + n_L\cos\theta_L} \tag{8.8.6}$$

The trace of F_T is as in Eq. (6.3.13):

$$\boxed{a = \frac{\cos(\delta_H + \delta_L) - \rho_T^2\cos(\delta_H - \delta_L)}{1 - \rho_T^2}} \tag{8.8.7}$$

The eigenvalues of the matrix F_T are $\lambda_\pm = e^{\pm jKl}$, where $K = \mathrm{acos}(a)/l$ and $l = l_H + l_L$. The condition $a = -1$ determines the bandedge frequencies of the high-reflectance bands. As in Eq. (6.3.16), this condition is equivalent to:

$$\boxed{\cos^2\Big(\frac{\delta_H + \delta_L}{2}\Big) = \rho_T^2\cos^2\Big(\frac{\delta_H - \delta_L}{2}\Big)} \tag{8.8.8}$$

Defining the quantities $L_\pm = L_H\cos\theta_H \pm L_L\cos\theta_L$ and the normalized frequency $F = f/f_0$, we may write:

$$\frac{\delta_H \pm \delta_L}{2} = \pi\frac{f}{f_0}(L_H\cos\theta_H \pm L_L\cos\theta_L) = \pi F L_\pm \tag{8.8.9}$$

Then, taking square roots of Eq. (8.8.8), we have:

$$\cos(\pi F L_+) = \pm|\rho_T|\cos(\pi F L_-)$$

The plus sign gives the left bandedge, $F_1 = f_1/f_0$, and the minus sign, the right bandedge, $F_2 = f_2/f_0$. Thus, F_1, F_2 are the solutions of the equations:

$$\boxed{\begin{aligned} \cos(\pi F_1 L_+) &= |\rho_T|\cos(\pi F_1 L_-) \\ \cos(\pi F_2 L_+) &= -|\rho_T|\cos(\pi F_2 L_-) \end{aligned}} \tag{8.8.10}$$

The bandwidth and center frequency of the reflecting band are:

$$\boxed{\frac{\Delta f}{f_0} = \Delta F = F_2 - F_1\,,\qquad \frac{f_c}{f_0} = F_c = \frac{F_1 + F_2}{2}} \tag{8.8.11}$$

The corresponding bandwidth in wavelengths is defined in terms of the left and right bandedge wavelengths:

$$\boxed{\lambda_1 = \frac{\lambda_0}{F_2} = \frac{c_0}{f_2}\,,\qquad \lambda_2 = \frac{\lambda_0}{F_1} = \frac{c_0}{f_1}\,,\qquad \Delta\lambda = \lambda_2 - \lambda_1} \tag{8.8.12}$$

An approximate solution of Eq. (8.8.10) can be obtained by setting $L_- = 0$ in the right-hand sides of Eq. (8.8.10):

$$\cos(\pi F_1 L_+) = |\rho_T|\,,\quad \cos(\pi F_2 L_+) = -|\rho_T| \tag{8.8.13}$$

with solutions:

$$F_1 = \frac{\mathrm{acos}(|\rho_T|)}{\pi L_+}\,,\quad F_2 = \frac{\mathrm{acos}(-|\rho_T|)}{\pi L_+} \tag{8.8.14}$$

Using the trigonometric identities $\mathrm{acos}(\pm|\rho_T|) = \pi/2 \mp \mathrm{asin}(|\rho_T|)$, we obtain the bandwidth and center frequency:

$$\boxed{\Delta f = f_2 - f_1 = \frac{2f_0\,\mathrm{asin}(|\rho_T|)}{\pi L_+}\,,\qquad f_c = \frac{f_1 + f_2}{2} = \frac{f_0}{2L_+}} \tag{8.8.15}$$

It follows that the center wavelength will be $\lambda_c = c_0/f_c = 2L_+\lambda_0$ or,

$$\boxed{\lambda_c = 2L_+\lambda_0 = 2\big(l_H n_H\cos\theta_H + l_L n_L\cos\theta_L\big)} \tag{8.8.16}$$

At normal incidence, we have $\lambda_c = 2(l_H n_H + l_L n_L)$. For quarter-wavelength designs at λ_0 at normal incidence, we have $L_+ = 1/4 + 1/4 = 1/2$, so that $\lambda_c = \lambda_0$.

The accuracy of the approximate solution (8.8.14) depends on the ratio $d = L_-/L_+$. Even if at normal incidence the layers were quarter-wavelength with $L_H = L_L = 0.25$, the equality of L_H and L_L will no longer be true at other angles of incidence. In fact, the quantity d is an increasing function of θ_a. For larger values of d, the exact solution of (8.8.10) can be obtained by the following iteration:

```
initialize with $F_1 = F_2 = 0$,
for $i = 0, 1, \dots, N_{\text{iter}}$, do:
    $F_1 = \frac{1}{\pi L_+}\,\text{acos}\big(|\rho_T|\cos(\pi F_1 L_-)\big)$
    $F_2 = \frac{1}{\pi L_+}\,\text{acos}\big(-|\rho_T|\cos(\pi F_2 L_-)\big)$
```
(8.8.17)

Evidently, the $i = 0$ iteration gives the zeroth-order solution (8.8.14). The iteration converges extremely fast, requiring only 3–4 iterations N_{iter}. The MATLAB function `omniband` implements this algorithm. It has usage:

```
[F1,F2] = omniband(na,nH,nL,LH,LL,theta,pol,Niter)     % bandedge frequencies
[F1,F2] = omniband(na,nH,nL,LH,LL,theta,pol)           % equivalent to Niter = 0
```

where `theta` is the incidence angle in degrees, `pol` is one of the strings `'te'` or `'tm'` for TE or TM polarization, and `Niter` is the desired number of iterations. If this argument is omitted, only the $i = 0$ iteration is carried out.

It is straightforward but tedious to verify the following facts about the above solutions. First, f_1, f_2 are *increasing* functions of θ_a for both TE and TM polarizations. Thus, the center frequency of the band $f_c = (f_1 + f_2)/2$ shifts towards higher frequencies with increasing angle θ_a. The corresponding wavelength intervals will shift towards lower wavelengths.

Second, the bandwidth $\Delta f = f_2 - f_1$ is an increasing function of θ_a for TE, and a decreasing one for TM polarization. Thus, as θ_a increases, the reflecting band for TE *expands* and that of TM *shrinks*, while their (slightly different) centers f_c shift upwards.

In order to achieve omnidirectional reflectivity, the TE and TM bands must have a common overlapping intersection for all angles of incidence. Because the TM band is always narrower than the TE band, it will determine the final common omnidirectional band.

The worst case of overlap is for the TM band at 90° angle of incidence, which must overlap with the TM/TE band at 0°. The left bandedge of this TM band, $f_{1,TM}(90^o)$, must be *less* than the right bandedge of the 0° band, $f_2(0^o)$. This is a sufficient condition for omnidirectional reflectivity.

Thus, the minimum band shared by all angles of incidence and both polarizations will be $[f_{1,TM}(90^o), f_2(0^o)]$, having width:

$$\boxed{\Delta f_{\min} = f_2(0^o) - f_{1,TM}(90^o)} \qquad \text{(minimum omnidirectional bandwidth)} \tag{8.8.18}$$

In a more restricted sense, the common reflecting band for both polarizations and for angles up to a given θ_a will be $[f_{1,TM}(\theta_a), f_{2,TM}(0^o)]$ and the corresponding bandwidth:

$$\Delta f(\theta_a) = f_2(0^o) - f_{1,TM}(\theta_a) \tag{8.8.19}$$

In addition to computing the bandwidths of either the TM or the TE bands at any angle of incidence, the function `omniband` can also compute the above common bandwidths. If the parameter `pol` is equal to `'tem'`, then F_1, F_2 are those of Eqs. (8.8.18) and (8.8.19). Its extended usage is as follows:

```
[F1,F2] = omniband(na,nH,nL,LH,LL,theta,'tem')       % Eq. (8.8.19)
[F1,F2] = omniband(na,nH,nL,LH,LL,90,'tem')          % Eq. (8.8.18)
[F1,F2] = omniband(na,nH,nL,LH,LL)                   % Eq. (8.8.18)
```

Next, we discuss some simulation examples that will help clarify the above remarks.

Example 8.8.1: The first example is the angular dependence of Example 6.3.2. In order to flatten out and sharpen the edges of the reflecting bands, we use $N = 30$ bilayers. Fig. 8.8.2 shows the TE and TM reflectances $|\Gamma_{T1}(\lambda)|^2$ as functions of the free-space wavelength λ, for the two angles of incidence $\theta_a = 45^o$ and 80^o.

Fig. 8.8.3 depicts the reflectances as functions of frequency f. The refractive indices were $n_a = 1$, $n_H = 2.32$, $n_L = 1.38$, $n_b = 1.52$, and the bilayers were quarter-wavelength $L_H = L_L = 0.25$ at the normalization wavelength $\lambda_0 = 500$ nm.

The necessary condition (8.8.3) is satisfied and we find for the maximum angle of refraction and the Brewster angle: $\theta_{H,\max} = 25.53^o$ and $\theta_B = 30.75^o$ Thus, we have $\theta_{H,\max} < \theta_B$.

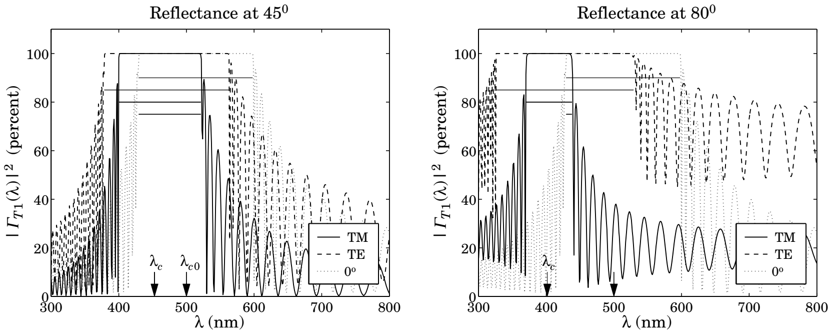


Fig. 8.8.2 TM and TE reflectances for $n_H = 2.32$, $n_L = 1.38$.

On each graph, we have indicated the corresponding bandwidth intervals calculated with `omniband`. The indicated intervals are for 0° incidence, for TE and TM, and for the common band Eq. (8.8.19) at θ_a. We observe the shifting of the bands towards higher frequencies,

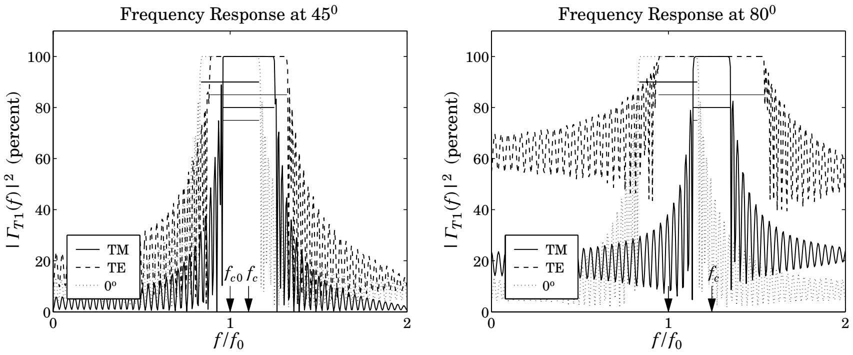


Fig. 8.8.3 TM and TE frequency responses for $n_H = 2.32$, $n_L = 1.38$.

or lower wavelengths, and the shrinking of the TM and expanding of the TE bands, and the shrinking of the common band.

At 45°, there is still sufficient overlap, but at 80°, the TM band has shifted almost to the end of the 0° band, resulting in an extremely narrow common band.

The arrows labeled f_{c0} and f_c represent the (TM) band center frequencies at 0° and 45° or 80°. The calculated bandedges corresponding to 90° incidence were $\lambda_1 = \lambda_0/F_{2,TM}(0^o) = 429.73$ nm and $\lambda_2 = \lambda_0/F_{1,TM}(90^o) = 432.16$ nm, with bandwidth $\Delta\lambda = \lambda_2 - \lambda_1 = 2.43$ nm. Thus, this structure does exhibit omnidirectional reflectivity, albeit over a very narrow band. The MATLAB code used to generate these graphs was:

```
na = 1; nb = 1.52; nH = 2.32; nL = 1.38;
LH = 0.25; LL = 0.25;

la0 = 500;
la = linspace(300,800,501);

th = 45; N = 30;
n = [na, nH, repmat([nL,nH], 1, N), nb];
L = [LH, repmat([LL,LH], 1, N)];
Ge = 100*abs(multidiel(n,L,la/la0, th, ’te’)).^2;
Gm = 100*abs(multidiel(n,L,la/la0, th, ’tm’)).^2;
G0 = 100*abs(multidiel(n,L,la/la0)).^2;

plot(la,Gm, la,Ge, la,G0);

[F10,F20] = omniband(na,nH,nL,LH,LL, 0, ’te’);
[F1e,F2e] = omniband(na,nH,nL,LH,LL, th,’te’);
[F1m,F2m] = omniband(na,nH,nL,LH,LL, th,’tm’);
[F1,F2] =   omniband(na,nH,nL,LH,LL, th,’tem’);
```

Because the reflectivity bands shrink with decreasing ratio n_H/n_L, if we were to slightly decrease n_H, then the TM band could be made to shift beyond the end of the 0° band and there would be no common overlapping reflecting band for all angles. We can observe this behavior in Fig. 8.8.4, which has $n_H = 2$, with all the other parameters kept the same.

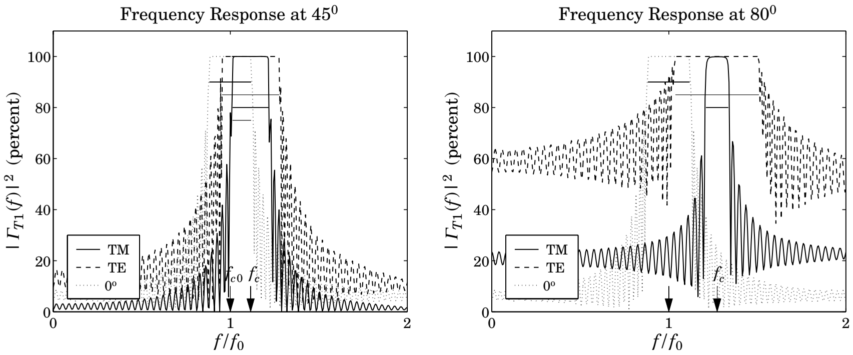


Fig. 8.8.4 TM and TE reflectances for $n_H = 2$, $n_L = 1.38$.

At 45° there is a common overlap, but at 80°, the TM band has already moved beyond the 0° band, while the TE band still overlaps with the latter. This example has no omnidirectional reflectivity, although the necessary condition (8.8.3) is still satisfied with $\theta_{H,\max} = 30^o$ and $\theta_B = 34.61^o$.

On the other hand, if we were to increase n_H, all the bands will widen, and so will the final common band, resulting in an omnidirectional mirror of wider bandwidth. Fig. 8.8.5 shows the case of $n_H = 3$, exhibiting a substantial overlap and omnidirectional behavior.

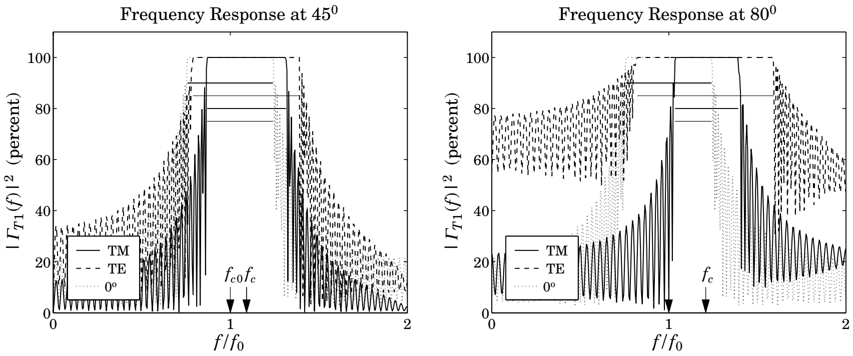


Fig. 8.8.5 TM and TE reflectances for $n_H = 3$, $n_L = 1.38$.

The minimum band (8.8.18) was $[F_1, F_2] = [1.0465, 1.2412]$ corresponding to the wavelength bandedges $\lambda_1 = \lambda_0/F_2 = 402.84$ nm and $\lambda_2 = \lambda_0/F_1 = 477.79$ nm with a width of $\Delta\lambda = \lambda_2 - \lambda_1 = 74.95$ nm, a substantial difference from that of Fig. 8.8.2. The bandedges were computed with $N_{\rm it} = 0$ in Eq. (8.8.17); with $N_{\rm it} = 3$, we obtain the more accurate values: $[F_1, F_2] = [1.0505, 1.2412]$.

To illustrate the dependence of the TE and TM bandwidths on the incident angle θ_a, we have calculated and plotted the normalized bandedge frequencies $F_1(\theta_a), F_2(\theta_a)$ for the range of angles $0 \le \theta_a \le 90^o$ for both polarizations. The left graph of Fig. 8.8.6 shows the case $n_H = 3$, $n_L = 1.38$, and the right graph, the case $n_H = 2$, $n_L = 1.38$.

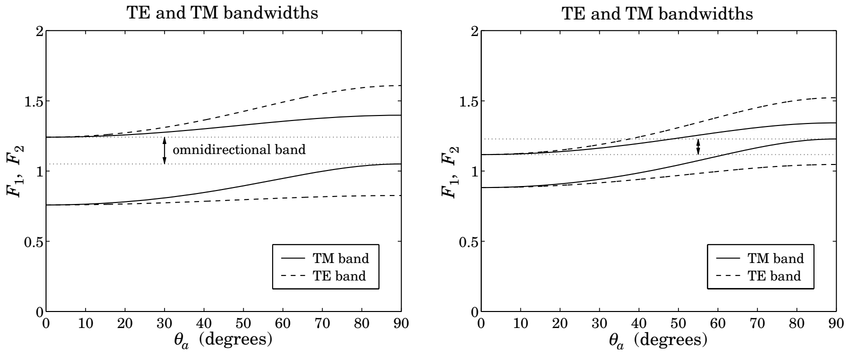


Fig. 8.8.6 TM/TE bandgaps versus angle for $n_H = 3$, $n_L = 1.38$ and $n_H = 2$, $n_L = 1.38$.

We note that the TE band widens with increasing angle, whereas the TM band narrows. At the same time, the band centers move toward higher frequencies. In the left graph, there is a common band shared by both polarizations and all angles, that is, the band defined by $F_2(0^o)$, and $F_{1,\mathrm{TM}}(90^o)$. For the right graph, the bandedge $F_{1,\mathrm{TM}}(\theta_a)$ increases beyond $F_2(0^o)$ for angles θ_a greater than about 61.8°, and therefore, there is no omnidirectional band. The calculations of $F_1(\theta_a)$, $F_2(\theta_a)$ were done with `omniband` with $N_{\mathrm{iter}} = 3$. □

Example 8.8.2: In Fig. 8.8.7, we study the effect of changing the optical lengths of the bilayers from quarter-wavelength to $L_H = 0.3$ and $L_L = 0.1$. The main result is to narrow the bands. This example, also illustrates the use of the iteration (8.8.17). The approximate solution (8.8.15) and exact solutions for the 80° bandedge frequencies are obtained from the two MATLAB calls:

```
[F1,F2] = omniband(na,nH,nL,LH,LL,80,'tem',0);
[F1,F2] = omniband(na,nH,nL,LH,LL,80,'tem',3);
```

with results $[F_1, F_2]=$ $[1.0933, 1.3891]$ and $[F_1, F_2]=$ $[1.1315, 1.3266]$, respectively. Three iterations produce an excellent approximation to the exact solution. □

Example 8.8.3: Here, we revisit Example 6.3.3, whose parameters correspond to the recently constructed omnidirectional infrared mirror [736]. Fig. 8.8.8 shows the reflectances as functions of wavelength and frequency at $\theta_a = 45^o$ and 80^o for both TE and TM polarizations. At both angles of incidence there is a wide overlap, essentially over the desired 10–15 μm band.

The structure consisted of nine alternating layers of Tellurium ($n_H = 4.6$) and Polystyrene ($n_L = 1.6$) on a NaCl substrate ($n_b = 1.48$.) The physical lengths were $l_H = 0.8$ and $l_L = 1.6$ μm. The normalizing wavelength was $\lambda_0 = 12.5$ μm. The optical thicknesses in units of λ_0 were $L_H = 0.2944$ and $L_L = 0.2112$.

The bandedges at 0° were $[F_1, F_2]= [0.6764, 1.2875]$ with center frequency $F_{c0} = 0.9819$, corresponding to wavelength $\lambda_{c0} = \lambda_0/F_{c0} = 12.73$ μm. Similarly, at 45°, the band centers for TE and TM polarizations were $F_{c,TE} = 1.0272$ and $F_{c,TM} = 1.0313$, resulting in the wavelengths $\lambda_{c,TE} = 12.17$ and $\lambda_{c,TM} = 12.12$ μm (shown on the graphs are the TE centers

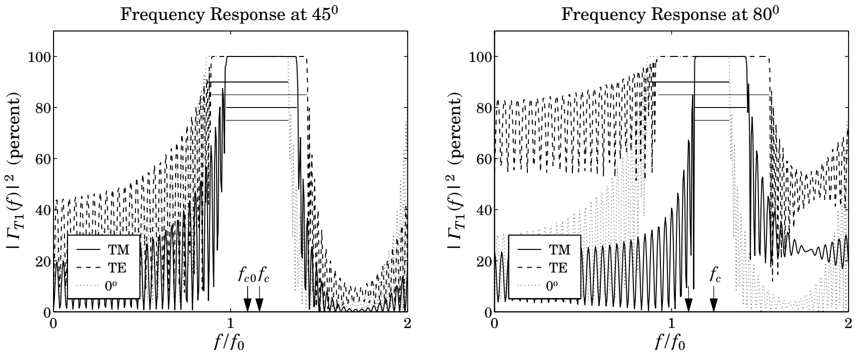


Fig. 8.8.7 Unequal length layers $L_H = 0.30$, $L_L = 0.15$.

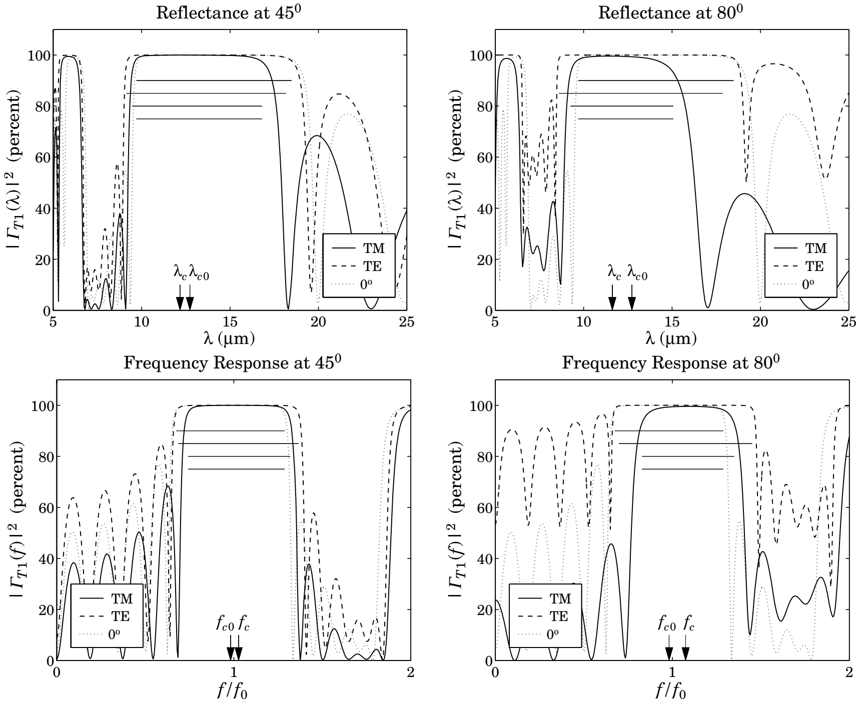


Fig. 8.8.8 Nine-layer Te/PS omnidirectional mirror over the infrared.

only.)

The final bandedges of the common reflecting band computed from Eq. (8.8.18) were $[F_1, F_2] = [0.8207, 1.2875]$, resulting in the wavelength bandedges $\lambda_1 = \lambda_0/F_2 = 9.71$ and $\lambda_2 = \lambda_0/F_1 = 14.95$ μm, with a width of $\Delta\lambda = \lambda_2 - \lambda_1 = 5.24$ μm and band center $(\lambda_1 + \lambda_2)/2 = 12.33$ μm (the approximation (8.8.15) gives 5.67 and 12.4 μm.) The graphs were generated by the following MATLAB code:

```
la0 = 12.5; la = linspace(5,25,401);
na = 1; nb = 1.48; nH = 4.6; nL = 1.6;
lH = 0.8; lL = 1.65;  LH = nH*lH/la0; LL = nL*lL/la0;

th = 45;
N = 4;
n = [na, nH, repmat([nL,nH], 1, N), nb];
L = [LH, repmat([LL,LH], 1, N)];
Ge = 100*abs(multidiel(n,L,la/la0, th, 'te')).^2;
Gm = 100*abs(multidiel(n,L,la/la0, th, 'tm')).^2;
G0 = 100*abs(multidiel(n,L,la/la0)).^2;

plot(la,Gm, la,Ge, la,G0);

Ni = 5;
[F10,F20] = omniband(na,nH,nL,LH,LL, 0, 'te', Ni);      band at 0°
[F1e,F2e] = omniband(na,nH,nL,LH,LL, th,'te', Ni);      TE band
[F1m,F2m] = omniband(na,nH,nL,LH,LL, th,'tm', Ni);      TM band
[F1,F2]   = omniband(na,nH,nL,LH,LL, th,'tem',Ni);      Eq. (8.8.19)
[F1,F2]   = omniband(na,nH,nL,LH,LL, 90,'tem',Ni);      Eq. (8.8.18)
```

Finally, Fig. 8.8.9 shows the same example with the number of bilayers doubled to $N = 8$. The mirror bands are flatter and sharper, but the widths are the same. □

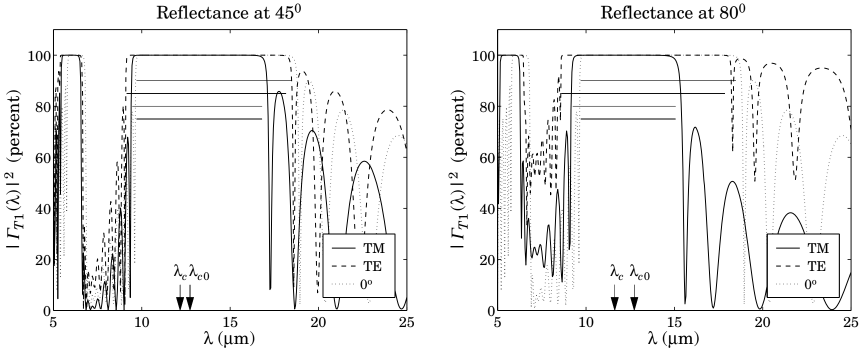


Fig. 8.8.9 Omnidirectional mirror with $N = 8$.

Example 8.8.4: The last example has parameters corresponding to the recently constructed omnidirectional reflector over the visible range [737]. The refractive indices were $n_a = 1$, $n_H = 2.6$ (ZnSe), $n_L = 1.34$ (Na_3AlF_6 cryolite), and $n_b = 1.5$ (glass substrate.) The layer lengths were $l_H = l_L = 90$ nm. There were $N = 9$ bilayers or $2N + 1 = 19$ layers, starting and ending with n_H.

With these values, the maximum angle of refraction is $\theta_{H,\max} = 22.27^o$ and is less than the Brewster angle $\theta_B = 27.27^o$.

The normalizing wavelength was taken to be $\lambda_0 = 620$ nm. Then, the corresponding optical lengths were $L_L = n_L l_L/\lambda_0 = 0.1945$ and $L_H = n_H l_H/\lambda_0 = 0.3774$. The overall minimum omnidirectional band is $[\lambda_1, \lambda_2] = [605.42,\ 646.88]$ nm. It was computed by the MATLAB call to `omniband` with $N_i = 5$ iterations:

```
[F1,F2] = omniband(na,nH,nL,LH,LL,90,'tem',Ni);
la1 = la0/F2; la2 = la0/F1;
```

(The values of λ_1, λ_2 do not depend on the choice of λ_0.) Fig. 8.8.10 shows the reflectance at 45^o and 80^o. The upper panel of graphs has $N = 9$ bilayers as in [737]. The lower panel has $N = 18$ bilayers or 38 layers, and has more well-defined band gaps. The two arrows in the figures correspond to the values of λ_1, λ_2 of the minimum omnidirectional band. □

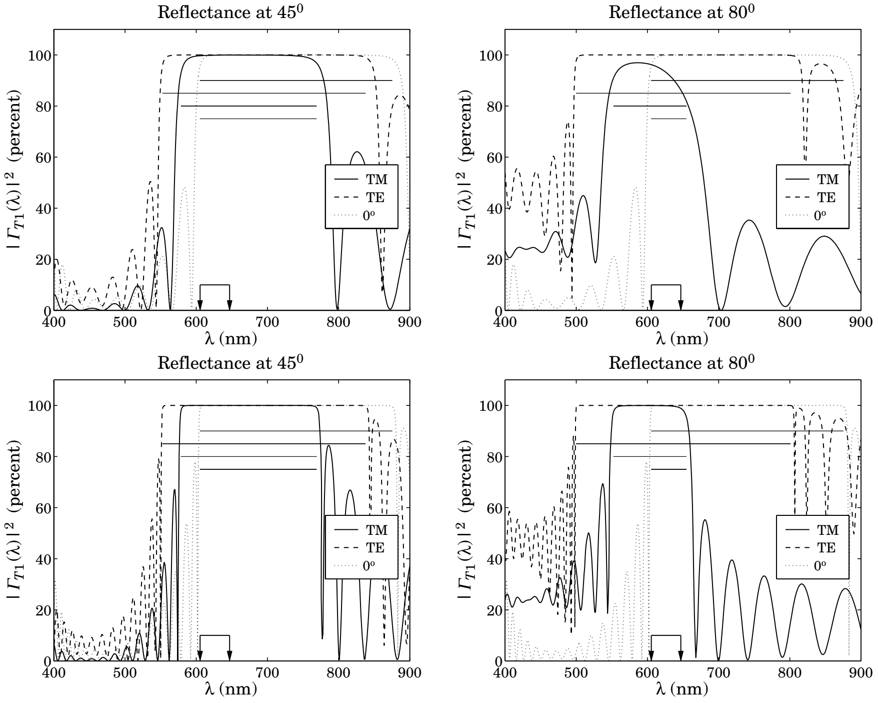


Fig. 8.8.10 Omnidirectional mirror over visible band.

8.9 Polarizing Beam Splitters

The objective of an omnidirectional mirror is to achieve high reflectivity for both polarizations. However, in polarizers, we are interested in separating the TE and TM polarizations. This can be accomplished with a periodic bilayer structure of the type shown in Fig. 8.8.1, which is highly reflecting only for TE and highly transmitting for TM polarizations. This is the principle of the so-called MacNeille polarizers [600,604,607,626,629,644–650].

If the angle of incidence θ_a is chosen such that the angle of refraction in the first high-index layer is equal to the Brewster angle of the high-low interface, then the TM component will not be reflected at the bilayer interfaces and will transmit through. The design condition is $\theta_H = \theta_B$, or $\sin\theta_H = \sin\theta_B$, which gives:

$$\boxed{n_a \sin\theta_a = n_H \sin\theta_H = n_H \sin\theta_B = \frac{n_H n_L}{\sqrt{n_H^2 + n_L^2}}} \tag{8.9.1}$$

This condition can be solved either for the angle θ_a or for the index n_a of the incident medium:

$$\boxed{\sin\theta_a = \frac{n_H n_L}{n_a\sqrt{n_H^2 + n_L^2}}} \quad \text{or,} \quad \boxed{n_a = \frac{n_H n_L}{\sin\theta_a\sqrt{n_H^2 + n_L^2}}} \tag{8.9.2}$$

In either case, the feasibility of this approach requires the opposite of the condition (8.8.3), that is,

$$n_a > \frac{n_H n_L}{\sqrt{n_H^2 + n_L^2}} \tag{8.9.3}$$

If the angle θ_a is set equal to the convenient value of 45°, then, condition Eq. (8.9.2) fixes the value of the refractive index n_a to be given by:

$$\boxed{n_a = \frac{\sqrt{2}\,n_H n_L}{\sqrt{n_H^2 + n_L^2}}} \tag{8.9.4}$$

Fig. 8.9.1 depicts such a multilayer structure sandwiched between two glass prisms with 45° angles. The thin films are deposited along the hypotenuse of each prism and the prisms are then cemented together. The incident, reflected, and transmitted beams are perpendicular to the prism sides.

Not many combinations of available materials satisfy condition (8.9.4). One possible solution is Banning's [607] with $n_H = 2.3$ (zinc sulfide), $n_L = 1.25$ (cryolite), and $n_a = 1.5532$. Another solution is given in Clapham, et al, [629], with $n_H = 2.04$ (zirconium oxide), $n_L = 1.385$ (magnesium fluoride), and $n_a = 1.6205$ (a form of dense flint glass.)

Fig. 8.9.2 shows the TE and TM reflectances of the case $n_H = 2.3$ and $n_L = 1.25$. The incident and output media had $n_a = n_b = 1.5532$. The maximum reflectivity for the TE component is 99.99%, while that of the TM component is 3% (note the different vertical scales in the two graphs.)

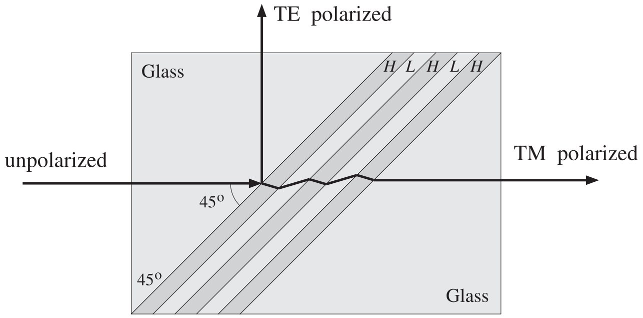


Fig. 8.9.1 Polarizing beam splitter.

The number of bilayers was $N = 5$ and the center frequency of the TE band was chosen to correspond to a wavelength of $\lambda_c = 500$ nm. To achieve this, the normalizing wavelength was required to be $\lambda_0 = 718.38$ nm. The layer lengths were quarter-wavelengths at λ_0. The TE bandwidth calculated with `omniband` is also shown.

The Brewster angles inside the high- and low-index layers are $\theta_H = 28.52°$ and $\theta_L = 61.48°$. As expected, they satisfy $\theta_H + \theta_L = 90°$.

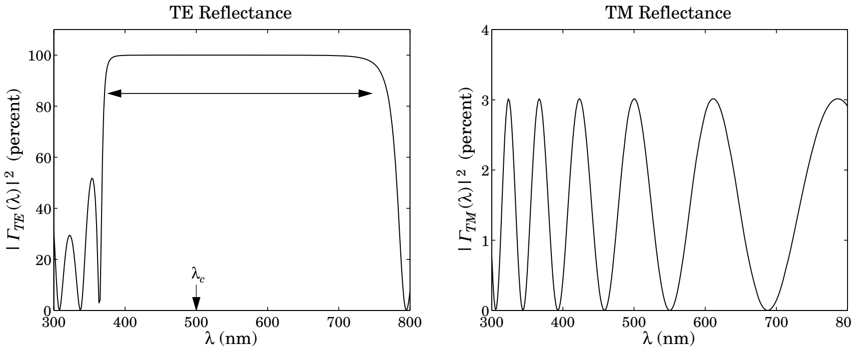


Fig. 8.9.2 Polarizer with $n_H = 2.3$ and $n_L = 1.25$.

Fig. 8.9.3 shows the second case having $n_H = 2.04$, $n_L = 1.385$, $n_a = n_b = 1.6205$. The normalizing wavelength was $\lambda_0 = 716.27$ nm in order to give $\lambda_c = 500$ nm. This case achieves a maximum TE reflectivity of 99.89% and TM reflectivity of only 0.53%. The typical MATLAB code generating these examples was:

```
nH = 2.3; nL = 1.25;
LH = 0.25; LL = 0.25;

na = nH*nL/sqrt(nH^2+nL^2)/sin(pi/4); nb=na;

[f1e,f2e] = omniband(na,nH,nL,LH,LL,th,’te’,5);
```

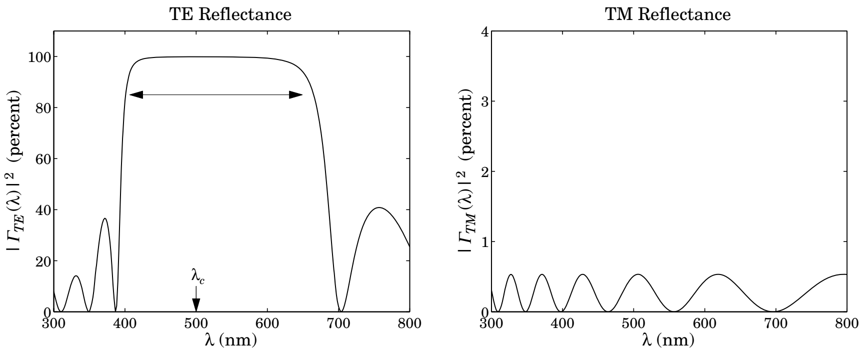


Fig. 8.9.3 Polarizer with $n_H = 2.04$ and $n_L = 1.385$.

```
lac = 500;
la0 = lac*(f1e+f2e)/2;                          because λc = λ0/Fc

la = linspace(300,800,301);

N = 5;
n = [na, nH, repmat([nL,nH], 1, N), nb];
L = [LH, repmat([LL,LH], 1, N)];
Ge = 100*abs(multidiel(n,L,la/la0, th, 'te')).^2;
Gm = 100*abs(multidiel(n,L,la/la0, th, 'tm')).^2;

plot(la,Ge);
```

8.10 Reflection and Refraction in Birefringent Media

Uniform plane wave propagation in biaxial media was discussed in Sec. 4.6. We found that there is an effective refractive index N such that $k = Nk_0 = N\omega/c_0$. The index N, given by Eq. (4.6.8), depends on the polarization of the fields and the direction of the wave vector. The expressions for the TE and TM fields were given in Eqs. (4.6.18) and (4.6.27).

Here, we discuss how such fields get reflected and refracted at planar interfaces between biaxial media. Further discussion can be found in [598] and [658–678].

Fig. 7.1.1 depicts the TM and TE cases, with the understanding that the left and right biaxial media are described by the triplets of principal indices $\mathbf{n} = [n_1, n_2, n_3]$ and $\mathbf{n}' = [n_1', n_2', n_3']$, and that the $\boldsymbol{E}$-fields are not perpendicular to the corresponding wave vectors in the TM case. The principal indices are aligned along the xyz axes, the xy-plane is the interface plane, and the xz-plane is the plane of incidence.

The boundary conditions require the matching of the electric field components that are tangential to the interface, that is, the components E_x in the TM case or E_y in TE. It proves convenient, therefore, to re-express Eq. (4.6.27) directly in terms of the E_x component and Eq. (4.6.18) in terms of E_y.

For the TM case, we write $\boldsymbol{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{z}}E_z = E_x(\hat{\mathbf{x}} - \hat{\mathbf{z}}\tan\bar{\theta})$, for the electric field of the left-incident field, where we used $E_z = -E_x\tan\bar{\theta}$. Similarly, for the magnetic field we have from Eq. (4.6.26):

$$\begin{aligned}
\boldsymbol{H} &= \frac{N}{\eta_0}\hat{\mathbf{y}}(E_x\cos\theta - E_z\sin\theta) = \frac{N}{\eta_0}\hat{\mathbf{y}}E_x\cos\theta\left(1 - \frac{E_z}{E_x}\tan\theta\right)\\
&= \frac{N}{\eta_0}\hat{\mathbf{y}}E_x\cos\theta\left(1 + \frac{n_1^2}{n_3^2}\tan^2\theta\right) = \frac{N}{\eta_0}\hat{\mathbf{y}}E_x\cos\theta\left(\frac{n_3^2\cos^2\theta + n_1^2\sin^2\theta}{n_3^2\cos^2\theta}\right)\\
&= \frac{N}{\eta_0}\hat{\mathbf{y}}E_x\cos\theta\left(\frac{n_3^2 n_1^2}{N^2 n_3^2\cos^2\theta}\right) = \frac{E_x}{\eta_0}\,\frac{n_1^2}{N\cos\theta}\,\hat{\mathbf{y}}
\end{aligned}$$

where we replaced $E_z/E_x = -\tan\bar{\theta} = -(n_1^2/n_3^2)\tan\theta$ and used Eq. (4.7.10). Thus,

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= E_x\left(\hat{\mathbf{x}} - \hat{\mathbf{z}}\,\frac{n_1^2}{n_3^2}\tan\theta\right)e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}\\
\boldsymbol{H}(\mathbf{r}) &= \frac{E_x}{\eta_0}\,\frac{n_1^2}{N\cos\theta}\,\hat{\mathbf{y}}\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}} = \frac{E_x}{\eta_{TM}}\,\hat{\mathbf{y}}\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}
\end{aligned}} \quad \text{(TM)} \qquad (8.10.1)$$

Similarly, we may rewrite the TE case of Eq. (4.6.18) in the form:

$$\boxed{\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= E_y\hat{\mathbf{y}}\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}\\
\boldsymbol{H}(\mathbf{r}) &= \frac{E_y}{\eta_0}\,n_2\cos\theta\,(-\hat{\mathbf{x}} + \hat{\mathbf{z}}\tan\theta)\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}} = \frac{E_y}{\eta_{TE}}\,(-\hat{\mathbf{x}} + \hat{\mathbf{z}}\tan\theta)\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}
\end{aligned}} \quad \text{(TE)} \quad (8.10.2)$$

The propagation phase factors are:

$$\boxed{e^{-j\,\boldsymbol{k}\cdot\mathbf{r}} = e^{-jk_0xN\sin\theta - jk_0zN\cos\theta}} \quad \text{(TM and TE propagation factors)} \qquad (8.10.3)$$

Unlike the isotropic case, the phase factors are different in the TM and TE cases because the value of N is different, as given by Eq. (4.6.8), or,

$$N = \begin{cases} \dfrac{n_1 n_3}{\sqrt{n_1^2\sin^2\theta + n_3^2\cos^2\theta}}\,, & \text{(TM or p-polarization)}\\[2ex] n_2, & \text{(TE or s-polarization)} \end{cases} \qquad (8.10.4)$$

In Eqs. (8.10.1) and (8.10.2), the effective *transverse impedances* are defined by $\eta_{TM} = E_x/H_y$ and $\eta_{TE} = -E_y/H_x$, and are given as follows:

$$\boxed{\eta_{TM} = \eta_0\,\frac{N\cos\theta}{n_1^2}\,, \quad \eta_{TE} = \frac{\eta_0}{n_2\cos\theta}} \quad \text{(transverse impedances)} \qquad (8.10.5)$$

Defining the TM and TE effective *transverse refractive indices* through $\eta_{TM} = \eta_0/n_{TM}$ and $\eta_{TE} = \eta_0/n_{TE}$, we have:

$$\boxed{\begin{aligned} n_{TM} &= \frac{n_1^2}{N\cos\theta} = \frac{n_1 n_3}{\sqrt{n_3^2 - N^2\sin^2\theta}} \\ n_{TE} &= n_2\cos\theta \end{aligned}} \qquad \text{(transverse refractive indices)} \tag{8.10.6}$$

where we used Eq. (4.6.23) for the TM case, that is,

$$\boxed{N\cos\theta = \frac{n_1}{n_3}\sqrt{n_3^2 - N^2\sin^2\theta}} \tag{8.10.7}$$

In the isotropic case, $N = n_1 = n_2 = n_3 = n$, Eqs. (8.10.6) reduce to Eq. (7.2.13). Next, we discuss the TM and TE reflection and refraction problems of Fig. 7.1.1.

Assuming that the interface is at $z = 0$, the equality of the total tangential electric fields (E_x component for TM and E_y for TE), implies as in Sec. 7.1 that the propagation phase factors must match at all values of x:

$$e^{-jk_{x+}x} = e^{-jk_{x-}x} = e^{-jk'_{x+}x} = e^{-jk'_{x-}x}$$

which requires that $k_{x+} = k_{x-} = k'_{x+} = k'_{x-}$, or, because $k_x = k\sin\theta = Nk_0\sin\theta$:

$$N\sin\theta_+ = N\sin\theta_- = N'\sin\theta'_+ = N'\sin\theta'_-$$

This implies Snel's law of reflection, that is, $\theta_+ = \theta_- \equiv \theta$ and $\theta'_+ = \theta'_- \equiv \theta'$, and Snel's law of refraction,

$$\boxed{N\sin\theta = N'\sin\theta'} \qquad \text{(Snel's law for birefringent media)} \tag{8.10.8}$$

Thus, Snel's law is essentially the same as in the isotropic case, provided one uses the effective refractive index N. Because N depends on the polarization, there will be two different refraction angles[†] for the same angle of incidence. In particular, Eq. (8.10.8) can be written explicitly in the two polarization cases:

$$\frac{n_1 n_3\sin\theta}{\sqrt{n_1^2\sin^2\theta + n_3^2\cos^2\theta}} = \frac{n'_1 n'_3\sin\theta'}{\sqrt{n_1'^2\sin^2\theta' + n_3'^2\cos^2\theta'}} \qquad \text{(TM)} \tag{8.10.9a}$$

$$n_2\sin\theta = n'_2\sin\theta' \qquad \text{(TE)} \tag{8.10.9b}$$

Both expressions reduce to Eq. (7.1.6) in the isotropic case. The explicit solutions of Eq. (8.10.9a) for $\sin\theta'$ and $\sin\theta$ are:

$$\begin{aligned} \sin\theta' &= \frac{n_1 n_3 n'_3\sin\theta}{\sqrt{\big[n_1'^2 n_3'^2(n_1^2 - n_3^2) - n_1^2 n_3^2(n_1'^2 - n_3'^2)\big]\sin^2\theta + n_1'^2 n_3'^2 n_3^2}} \\ \sin\theta &= \frac{n'_1 n'_3 n_3\sin\theta'}{\sqrt{\big[n_1^2 n_3^2(n_1'^2 - n_3'^2) - n_1'^2 n_3'^2(n_1^2 - n_3^2)\big]\sin^2\theta' + n_1^2 n_3^2 n_3'^2}} \end{aligned} \tag{8.10.10}$$

[†] Hence, the name birefringent.

The MATLAB function `Snel`, solves Eqs. (8.10.9) for θ' given the angle of incidence θ and the polarization type. It works for any type of medium, isotropic, uniaxial, or biaxial. It has usage:

```
thb = snel(na,nb,tha,pol);              % refraction angle from Snel's law
```

The refractive index inputs `na`, `nb` may be entered as 1-, 2-, or 3-dimensional column or row vectors, for example, $\mathbf{n}_a = [n_a]$ (isotropic), $\mathbf{n}_a = [n_{ao}, n_{ae}]$ (uniaxial), or $\mathbf{n}_a = [n_{a1}, n_{a2}, n_{a3}]$ (biaxial).

Next, we discuss the propagation and matching of the transverse fields. All the results of Sec. 7.3 translate verbatim to the birefringent case, provided one uses the proper transverse refractive indices according to Eq. (8.10.6).

In particular, the propagation equations (7.3.5)–(7.3.7) for the transverse fields, for the transverse reflection coefficients Γ_T, and for the transverse wave impedances Z_T, remain unchanged.

The *phase thickness* δ_z for propagating along z by a distance l also remains the same as Eq. (7.3.8), except that the index N must be used in the optical length, and therefore, δ_z depends on the polarization:

$$\boxed{\delta_z = k_z l = kl\cos\theta = Nk_0 l\cos\theta = \frac{2\pi}{\lambda}\,lN\cos\theta} \tag{8.10.11}$$

Using Eq. (8.10.7), we have explicitly:

$$\delta_z = \frac{2\pi}{\lambda}\,l\,\frac{n_1}{n_3}\sqrt{n_3^2 - N^2\sin^2\theta}\,, \qquad \text{(TM)} \tag{8.10.12a}$$

$$\delta_z = \frac{2\pi}{\lambda}\,ln_2\cos\theta\,, \qquad \text{(TE)} \tag{8.10.12b}$$

The transverse matching matrix (7.3.11) and Fresnel reflection coefficients (7.3.12) remain the same. Explicitly, we have in the TM and TE cases:

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{n_{TM} - n'_{TM}}{n_{TM} + n'_{TM}} = \frac{\dfrac{n_1^2}{N\cos\theta} - \dfrac{n_1'^2}{N'\cos\theta'}}{\dfrac{n_1^2}{N\cos\theta} + \dfrac{n_1'^2}{N'\cos\theta'}} \\ \rho_{TE} &= \frac{n_{TE} - n'_{TE}}{n_{TE} + n'_{TE}} = \frac{n_2\cos\theta - n'_2\cos\theta'}{n_2\cos\theta + n'_2\cos\theta'} \end{aligned}} \tag{8.10.13}$$

Using Eq. (8.10.6) and the TM and TE Snel's laws, Eqs. (8.10.9), we may rewrite the reflection coefficients in terms of the angle θ only:

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{n_1 n_3\sqrt{n_3'^2 - N^2\sin^2\theta} - n'_1 n'_3\sqrt{n_3^2 - N^2\sin^2\theta}}{n_1 n_3\sqrt{n_3'^2 - N^2\sin^2\theta} + n'_1 n'_3\sqrt{n_3^2 - N^2\sin^2\theta}} \\ \rho_{TE} &= \frac{n_2\cos\theta - \sqrt{n_2'^2 - n_2^2\sin^2\theta}}{n_2\cos\theta + \sqrt{n_2'^2 - n_2^2\sin^2\theta}} \end{aligned}} \tag{8.10.14}$$

The quantity $N^2\sin^2\theta$ can be expressed directly in terms of θ and the refractive indices of the incident medium. Using Eq. (8.10.4), we have:

$$\boxed{N^2\sin^2\theta = \frac{n_1^2 n_3^2\sin^2\theta}{n_1^2\sin^2\theta + n_3^2\cos^2\theta}} \tag{8.10.15}$$

The TE reflection coefficient behaves like the TE isotropic case. The TM coefficient exhibits a much more complicated behavior. If $n_1 = n_1'$ but $n_3 \neq n_3'$, it behaves like the TM isotropic case. If $n_3 = n_3'$ but $n_1 \neq n_1'$, the square-root factors cancel and it becomes independent of θ:

$$\boxed{\rho_{TM} = \frac{n_1 - n_1'}{n_1 + n_1'}} \tag{8.10.16}$$

Another interesting case is when both media are uniaxial and $n_3' = n_1$ and $n_1' = n_3$, that is, the refractive index vectors are $\mathbf{n} = [n_1, n_1, n_3]$ and $\mathbf{n}' = [n_3, n_3, n_1]$. It is straightforward to show in this case that $\rho_{TM} = \rho_{TE}$ at all angles of incidence. Multilayer films made from alternating such materials exhibit similar TM and TE optical properties [658].

The MATLAB function `fresnel` can evaluate Eqs. (8.10.14) at any range of incident angles θ. The function determines internally whether the media are isotropic, uniaxial, or biaxial.

8.11 Brewster and Critical Angles in Birefringent Media

The maximum angle of refraction, critical angle of incidence, and Brewster angle, have their counterparts in birefringent media.

It is straightforward to verify that θ' is an increasing function of θ in Eq. (8.10.9). The maximum angle of refraction θ_c' is obtained by setting $\theta = 90^o$ in Eq. (8.10.9).

For the TE case, we obtain $\sin\theta_c' = n_2/n_2'$. As in the isotropic case of Eq. (7.5.2), this requires that $n_2 < n_2'$, that is, the incident medium is less dense than the transmitted medium, with respect to the index n_2. For the TM case, we obtain from Eq. (8.10.9a):

$$\boxed{\sin\theta_c' = \frac{n_3 n_3'}{\sqrt{n_3^2 n_3'^2 + n_1'^2(n_3'^2 - n_3^2)}}} \qquad \text{(maximum TM refraction angle)} \tag{8.11.1}$$

This requires that $n_3 < n_3'$. On the other hand, if $n_3 > n_3'$, we obtain the critical angle of incidence θ_c that corresponds to $\theta' = 90^o$ in Eq. (8.10.10):

$$\boxed{\sin\theta_c = \frac{n_3 n_3'}{\sqrt{n_3^2 n_3'^2 + n_1^2(n_3^2 - n_3'^2)}}} \qquad \text{(critical TM angle)} \tag{8.11.2}$$

whereas for the TE case, we have $\sin\theta_c = n_2'/n_2$, which requires $n_2 > n_2'$.

In the isotropic case, a Brewster angle always exists at which the TM reflection coefficient vanishes, $\rho_{TM} = 0$. In the birefringent case, the Brewster angle does not necessarily exist, as is the case of Eq. (8.10.16), and it can also have the value zero, or even be imaginary.

The Brewster angle condition $\rho_{TM} = 0$ is equivalent to the equality of the transverse refractive indices $n_{TM} = n_{TM}'$. Using Eq. (8.10.6), we obtain:

$$n_{TM} = n_{TM}' \quad\Rightarrow\quad \frac{n_1 n_3}{\sqrt{n_3^2 - N^2\sin^2\theta}} = \frac{n_1' n_3'}{\sqrt{n_3'^2 - N^2\sin^2\theta}} \tag{8.11.3}$$

where $N^2\sin^2\theta$ is given by Eq. (8.10.15). Solving for θ, we obtain the expression for the Brewster angle from the left medium:

$$\boxed{\tan\theta_B = \frac{n_3 n_3'}{n_1^2}\sqrt{\frac{n_1^2 - n_1'^2}{n_3^2 - n_3'^2}}} \qquad \text{(Brewster angle)} \tag{8.11.4}$$

Working instead with $N'\sin\theta' = N\sin\theta$, we obtain the Brewster angle from the right medium, interchanging the roles of the primed and unprimed quantities:

$$\boxed{\tan\theta_B' = \frac{n_3 n_3'}{n_1'^2}\sqrt{\frac{n_1^2 - n_1'^2}{n_3^2 - n_3'^2}}} \qquad \text{(Brewster angle)} \tag{8.11.5}$$

Eqs. (8.11.4) and (8.11.5) reduce to Eqs. (7.6.2) and (7.6.3) in the isotropic case. It is evident from Eq. (8.11.4) that θ_B is a real angle only if the quantity under the square root is non-negative, that is, only if $n_1 > n_1'$ and $n_3 > n_3'$, or if $n_1 < n_1'$ and $n_3 < n_3'$. Otherwise, θ_B is imaginary. In the special case, $n_1 = n_1'$ but $n_3 \neq n_3'$, the Brewster angle vanishes. If $n_3 = n_3'$, the Brewster angle does not exist, since then ρ_{TM} is given by Eq. (8.10.16) and cannot vanish.

The MATLAB function `brewster` computes the Brewster angle θ_B, as well as the critical angles θ_c and θ_c'. For birefringent media the critical angles depend on the polarization. Its usage is as follows:

```
      [thB,thc] = brewster(na,nb)         % isotropic case
[thB,thcTE,thcTM] = brewster(na,nb)       % birefringent case
```

In multilayer systems, it is convenient to know if the Brewster angle of an internal interface is accessible from the incident medium. Using Snel's law we have in this case $N_a\sin\theta_a = N\sin\theta$, where θ_a is the incident angle and N_a the effective index of the incident medium. It is simpler, then, to solve Eq. (8.11.3) directly for θ_a:

$$\boxed{N_a^2\sin\theta_a^2 = N^2\sin^2\theta_B = \frac{n_3^2 n_3'^2(n_1^2 - n_1'^2)}{n_1^2 n_3^2 - n_3'^2 n_1'^2}} \tag{8.11.6}$$

Example 8.11.1: To illustrate the variety of possible Brewster angle values, we consider the following birefringent cases:

(a) $\mathbf{n} = [1.63, 1.63, 1.5]$, $\mathbf{n}' = [1.63, 1.63, 1.63]$
(b) $\mathbf{n} = [1.54, 1.54, 1.63]$, $\mathbf{n}' = [1.5, 1.5, 1.5]$
(c) $\mathbf{n} = [1.8, 1.8, 1.5]$, $\mathbf{n}' = [1.5, 1.5, 1.5]$
(d) $\mathbf{n} = [1.8, 1.8, 1.5]$, $\mathbf{n}' = [1.56, 1.56, 1.56]$

These cases were discussed in [658]. The corresponding materials are: (a) birefringent polyester and isotropic polyester, (b) syndiotactic polystyrene and polymethylmethacrylate (PMMA), (c) birefringent polyester and PMMA, and (d) birefringent polyester and isotropic polyester.

Because $n_1 = n_1'$ in case (a), the Brewster angle will be zero, $\theta_B = 0^\circ$. In case (b), we calculate $\theta_B = 29.4^\circ$. Because $n_2 > n_2'$ and $n_3 > n_3'$, there will be both TE and TM critical angles of reflection: $\theta_{c,TE} = 76.9^\circ$ and $\theta_{c,TM} = 68.1^\circ$.

In case (c), the Brewster angle does not exist because $n_3 = n_3'$, and in fact, the TM reflection coefficient is independent of the incident angle as in Eq. (8.10.16). The corresponding critical angles of reflection are: $\theta_{c,TE} = 56.4^\circ$ and $\theta_{c,TM} = 90^\circ$.

Finally, in case (d), because $n_2 > n_2'$ but $n_3 < n_3'$, the Brewster angle will be imaginary, and there will be a TE critical angle of reflection and a TM maximum angle of refraction: $\theta_{c,TE} = 60.1^\circ$ and $\theta'_{c,TM} = 74.1^\circ$.

Fig. 8.11.1 shows the TM and TE reflection coefficients $|\rho_{TM}(\theta)|$ of Eq. (8.10.14) versus θ in the range $0 \le \theta \le 90^\circ$.

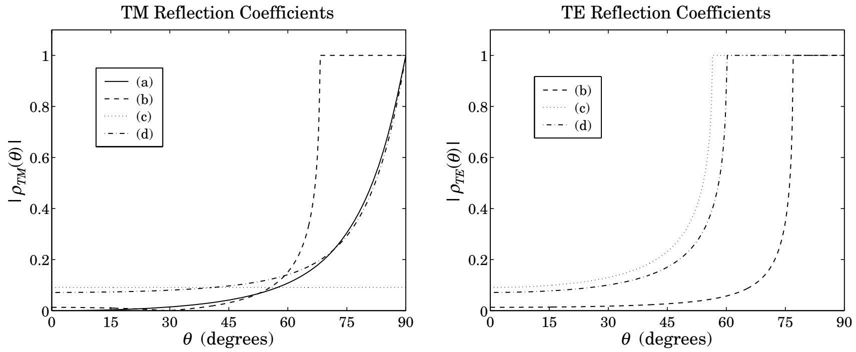


Fig. 8.11.1 TM and TE birefringent Fresnel reflection coefficients versus incident angle.

The TE coefficient in case (a) is not plotted because it is identically zero. In order to expand the vertical scales, Fig. 8.11.2 shows the TM reflectances normalized by their values at $\theta = 0^\circ$, that is, it plots the quantities $|\rho_{TM}(\theta)/\rho_{TM}(0^\circ)|^2$. Because in case (a) $\rho_{TM}(0^\circ) = 0$, we have plotted instead the scaled-up quantity $|100\rho_{TM}(\theta)|^2$.

The typical MATLAB code used to compute the critical angles and generate these graphs was:

```
th = linspace(0,90,361);                         % θ at 1/4° intervals

na = [1.63,1.63,1.5]; nb = [1.63,1.63,1.63];     % note the variety of
[rte1,rtm1] = fresnel(na,nb,th);                 % equivalent ways of
[thb1,thcTE1,thcTM1] = brewster(na,nb);          % entering na and nb

na = [1.54,1.63];
nb = [1.5, 1.5];                               % FRESNEL and BREWSTER
[rte2,rtm2] = fresnel(na,nb,th);               % internally extend
[thb2,thcTE2,thcTM2] = brewster(na,nb);        % na,nb into 3-d arrays
```

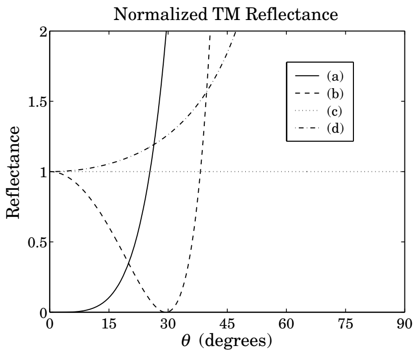


Fig. 8.11.2 TM reflectances normalized at normal incidence.

```
na = [1.8, 1.5];                               % same as na=[1.8,1.8,1.5]
nb = 1.5;                                      % and nb=[1.5,1.5,1.5]
[rte3,rtm3] = fresnel(na,nb,th);
[thb3,thcTE3,thcTM3] = brewster(na,nb);        % in this case, θB = []

na = [1.8,1.5];
nb = 1.56;
[rte4,rtm4] = fresnel(na,nb,th);
[thb4,thcTE4,thcTM4] = brewster(na,nb);

plot(th, abs([rtm1; rtm2; rtm3; rtm4]));
```

We note four striking properties of the birefringent cases that have no counterparts for isotropic materials: (i) The Brewster angle can be zero, (ii) the Brewster angle may not exist, (iii) the Brewster angle may be imaginary with the TE and TM reflection coefficients both increasing monotonically with the incident angle, and (iv) there may be total internal reflection in one polarization but not in the other.

8.12 Multilayer Birefringent Structures

With some redefinitions, all the results of Sec. 8.1 on multilayer dielectric structures translate essentially unchanged to the birefringent case.

We assume the same M-layer configuration shown in Fig. 8.1.1, where now each layer is a biaxial material. The orthogonal optic axes of all the layers are assumed to be aligned with the xyz film axes. The xz-plane is the plane of incidence, the layer interfaces are parallel to the xy-plane, and the layers are arranged along the z-axis.

The ith layer is described by the triplet of refractive indices $\mathbf{n}_i = [n_{i1}, n_{i2}, n_{i3}]$, $i = 1, 2, \dots, M$. The incident and exit media a, b may also be birefringent with $\mathbf{n}_a = [n_{a1}, n_{a2}, n_{a3}]$ and $\mathbf{n}_b = [n_{b1}, n_{b2}, n_{b3}]$, although in our examples, we will assume that they are isotropic.

The reflection/refraction angles in each layer depend on the assumed polarization and are related to each other by the birefringent version of Snel's law, Eq. (8.10.8):

$$\boxed{N_a\sin\theta_a = N_i\sin\theta_i = N_b\sin\theta_b}\,,\quad i = 1,2\dots,M \tag{8.12.1}$$

where N_a, N_i, N_b are the effective refractive indices given by Eq. (8.10.4). The phase thickness of the ith layer depends on the polarization:

$$\delta_i = \frac{2\pi}{\lambda}\,l_i N_i\cos\theta_i = \begin{cases} \dfrac{2\pi}{\lambda}\,l_i n_{i1}\sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i3}^2}}\,, & \text{(TM)} \\[2ex] \dfrac{2\pi}{\lambda}\,l_i n_{i2}\sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i2}^2}}\,, & \text{(TE)} \end{cases} \tag{8.12.2}$$

where we used Eq. (8.10.7) and Snel's law to write in the TM and TE cases:

$$N_i\cos\theta_i = \begin{cases} \dfrac{n_{i1}}{n_{i3}}\sqrt{n_{i3}^2 - N_i^2\sin^2\theta_i} = n_{i1}\sqrt{1-\dfrac{N_i^2\sin^2\theta_i}{n_{i3}^2}} = n_{i1}\sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i3}^2}} \\[2ex] n_{i2}\cos\theta_i = n_{i2}\sqrt{1-\sin^2\theta_i} = n_{i2}\sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i2}^2}} \end{cases}$$

To use a unified notation for the TM and TE cases, we define the layer optical lengths at normal-incidence, normalized by a fixed free-space wavelength λ_0:

$$L_i = \begin{cases} \dfrac{l_i n_{i1}}{\lambda_0}\,, & \text{(TM)} \\[1ex] \dfrac{l_i n_{i2}}{\lambda_0}\,, & \text{(TE)} \end{cases}\,,\quad i = 1,2,\dots,M \tag{8.12.3}$$

We define also the cosine coefficients c_i, which represent $\cos\theta_i$ in the TE birefringent case and in the TM isotropic case:

$$c_i = \begin{cases} \sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i3}^2}}\,, & \text{(TM)} \\[2ex] \sqrt{1-\dfrac{N_a^2\sin^2\theta_a}{n_{i2}^2}}\,, & \text{(TE)} \end{cases}\,,\quad i = 1,2,\dots,M \tag{8.12.4}$$

At normal incidence the cosine factors are unity, $c_i = 1$. With these definitions, Eq. (8.12.2) can be written compactly in the form:

$$\boxed{\delta_i = 2\pi\,\frac{\lambda_0}{\lambda}\,L_i c_i = 2\pi\,\frac{f}{f_0}\,L_i c_i}\,,\quad i = 1,2,\dots,M \tag{8.12.5}$$

where λ is the operating free-space wavelength and $f = c_0/\lambda$, $f_0 = c_0/\lambda_0$. This is the birefringent version of Eq. (8.1.10). A typical design might use quarter-wave layers, $L_i = 1/4$, at λ_0 and at normal incidence.

The reflection coefficients ρ_{Ti} at the interfaces are given by Eq. (8.1.3), but now the transverse refractive indices are defined by the birefringent version of Eq. (8.1.4):

$$n_{Ti} = \begin{cases} \dfrac{n_{i1}^2}{N_i\cos\theta_i} = \dfrac{n_{i1}n_{i3}}{\sqrt{n_{i3}^2 - N_a^2\sin^2\theta_a}}\,, & \text{(TM)} \\[2ex] n_{i2}\cos\theta_i = \sqrt{n_{i2}^2 - N_a^2\sin^2\theta_a}\,, & \text{(TE)} \end{cases}\,,\quad i = a,1,2,\dots,M,b \tag{8.12.6}$$

With the above redefinitions, the propagation and matching equations (8.1.5)–(8.1.9) remain unchanged. The MATLAB function `multidiel` can also be used in the birefringent case to compute the frequency reflection response of a multilayer structure. Its usage is still:

```
[Gamma1,Z1] = multidiel(n,L,lambda,theta,pol);     % birefringent multilayer structure
```

where the input `n` is a $1\times(M+2)$ vector of refractive indices in the isotropic case, or a $3\times(M+2)$ matrix, where each column represents the triplet of birefringent indices of each medium. For uniaxial materials, `n` may be entered as a $2\times(M+2)$ matrix.

8.13 Giant Birefringent Optics

The results of Sec. 8.8 can be applied almost verbatim to the birefringent case. In Fig. 8.8.1, we assume that the high and low alternating layers are birefringent, described by the triplet indices $\mathbf{n}_H = [n_{H1}, n_{H2}, n_{H3}]$ and $\mathbf{n}_L = [n_{L1}, n_{L2}, n_{L3}]$. The entry and exit media may also be assumed to be birefringent. Then, Snel's laws give:

$$N_a\sin\theta_a = N_H\sin\theta_H = N_L\sin\theta_L = N_b\sin\theta_b \tag{8.13.1}$$

The phase thicknesses δ_H and δ_L within the high and low index layers are:

$$\boxed{\delta_H = 2\pi\,\frac{f}{f_0}\,L_H c_H\,,\quad \delta_L = 2\pi\,\frac{f}{f_0}\,L_L c_L} \tag{8.13.2}$$

where L_H, c_H and L_L, c_L are defined by Eqs. (8.12.3) and (8.12.4) for $i = H, L$. The effective transverse refractive indices within the high and low index layers are given by Eq. (8.12.6), again with $i = H, L$.

The alternating reflection coefficient ρ_T between the high/low interfaces is given by Eq. (8.10.14), with the quantity $N^2\sin^2\theta$ replaced by $N_a^2\sin^2\theta_a$ by Snel's law:

$$\boxed{\begin{aligned} \rho_{TM} &= \frac{n_{H1}n_{H3}\sqrt{n_{L3}^2 - N_a^2\sin^2\theta_a} - n_{L1}n_{L3}\sqrt{n_{H3}^2 - N_a^2\sin^2\theta_a}}{n_{H1}n_{H3}\sqrt{n_{L3}^2 - N_a^2\sin^2\theta_a} + n_{L1}n_{L3}\sqrt{n_{H3}^2 - N_a^2\sin^2\theta_a}} \\ \rho_{TE} &= \frac{\sqrt{n_{H2}^2 - N_a^2\sin^2\theta_a} - \sqrt{n_{L2}^2 - N_a^2\sin^2\theta_a}}{\sqrt{n_{H2}^2 - N_a^2\sin^2\theta_a} + \sqrt{n_{L2}^2 - N_a^2\sin^2\theta_a}} \end{aligned}} \tag{8.13.3}$$

The multilayer structure will exhibit reflection bands whose bandedges can be calculated from Eqs. (8.8.7)–(8.8.17), with the redefinition $L_\pm = L_H c_H \pm L_L c_L$. The MATLAB function `omniband2` calculates the bandedges. It has usage:

```
[F1,F2] = omniband2(na,nH,nL,LH,LL,th,pol,N);
```

where `pol` is one of the strings `'te'` or `'tm'` for TE or TM polarization, and `na`, `nH`, `nL` are 1-d, 2-d, or 3-d row or column vectors of birefringent refractive indices.

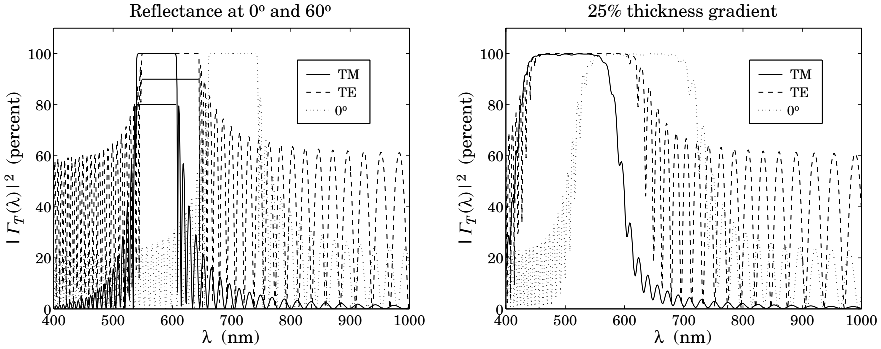
Next, we discuss some mirror design examples from [658] that illustrate some properties that are specific to birefringent media. The resulting optical effects in such mirror structures are referred to as *giant birefringent optics* (GBO) in [658,1329].

Example 8.13.1: We consider a GBO mirror consisting of 50-bilayers of high and low index quarter-wave layers with refractive indices $\mathbf{n}_H = [1.8, 1.8, 1.5]$, $\mathbf{n}_L = [1.5, 1.5, 1.5]$ (birefringent polyester and isotropic PMMA.) The surrounding media are air, $n_a = n_b = 1$.

The layers are quarter wavelength at the normalization wavelength $\lambda_0 = 700$ nm at normal incidence, so that for both polarizations we take $L_H = L_L = 1/4$.

Because the high/low index layers are matched along the z-direction, $n_{H3} = n_{L3}$, the TM reflection coefficient at the high/low interface will be constant, independent of the incident angle θ_a, as in Eq. (8.10.16). However, some dependence on θ_a is introduced through the cosine factors c_H, c_L of Eq. (8.13.2).

The left graph of Fig. 8.13.1 shows the reflectance $|\Gamma_T(\lambda)|^2$ as a function of λ for an angle of incidence $\theta_a = 60^\circ$. The TM and TE bandedge wavelengths were calculated from `omniband2` to be: $[\lambda_1, \lambda_2] = [540.24, 606.71]$ and $[\lambda_1, \lambda_2] = [548.55, 644.37]$ nm.


Reflectance at 0° and 60°

25% thickness gradient

Fig. 8.13.1 Reflectance of birefringent mirror.

The typical MATLAB code used to generate the left graph and the bandedge wavelengths was as follows:

```
LH = 0.25; LL = 0.25;

na = [1; 1; 1];
nH = [1.8; 1.8; 1.5];
nL = [1.5; 1.5; 1.5];
```

```
nb = [1; 1; 1];

la0 = 700;
la = linspace(400,1000,601);

th = 60;                                          % angle of incidence

N = 50;                                           % number of bilayers
n = [na, repmat([nH,nL], 1, N), nb];              % 3×(2N + 2) matrix
L = [repmat([LH,LL], 1, N)];

Ge = 100*abs(multidiel(n, L, la/la0, th, 'te')).^2;
Gm = 100*abs(multidiel(n, L, la/la0, th, 'tm')).^2;
G0 = 100*abs(multidiel(n, L, la/la0)).^2;

plot(la,Gm,'-', la,Ge,'--', la,G0,':');

[F1,F2]=omniband2(na,nH,nL,LH,LL,th,'tm',3);
la1 = la0/F2;  la2 = la0/F1;                      % TM bandedge wavelengths
```

The right graph shows the reflectance with a 25% thickness gradient (the layer thicknesses L_H, L_L decrease linearly from quarter-wavelength to 25% less than that at the end.) This can be implemented in MATLAB by defining the thickness vector L by:

```
L = [repmat([LH,LL], 1, N)];
L = L .* (1 - linspace(0, 0.25, 2*N));   % 25% thickness gradient
```

The thickness gradient increases the effective bandwidth of the reflecting bands [656]. However, the bandwidth calculation can no longer be done with `omniband2`. The band centers can be shifted to higher wavelengths by choosing λ_0 higher. The reflecting bands can be made flatter by increasing the number of bilayers. □

Example 8.13.2: In this example, we design a 30-bilayer GBO mirror with $\mathbf{n}_H = [1.8, 1.8, 1.5]$ and $\mathbf{n}_L = [1.5, 1.5, 1.8]$, so that $n_{H1} = n_{H2} = n_{L3}$ and $n_{H3} = n_{L1} = n_{L2}$. As we discussed in Sec. 8.10, it follows from Eq. (8.10.14) that $\rho_{TM} = \rho_{TE}$ for all angles of incidence.

As in Ref. [658], the media a, b are taken to be isotropic with $n_a = n_b = 1.4$. The normalization wavelength at which the high and low index layers are quarter-wavelength is $\lambda_0 = 700$ nm.

The left graph of Fig. 8.13.2 shows the reflectance for a 45° angle of incidence. Because $\rho_{TM} = \rho_{TE}$, the reflection bands for the TM and TE cases are essentially the same.

The right graph depicts the asymptotic (for large number of bilayers) bandedges of the reflecting band versus incident angle. They were computed with `omniband2`. Unlike the isotropic case, the TM and TE bands are exactly identical. This is a consequence of the following relationships between the cosine factors in this example: $c_{H,TM} = c_{L,TE}$ and $c_{H,TE} = c_{L,TM}$. Then, because we assume quarter-wave layers in both the TE and TM cases, $L_H = L_L = 1/4$, we will have:

$$L_{+,TM} = L_{H,TM}c_{H,TM} + L_{L,TM}c_{L,TM} == \frac{1}{4}(c_{H,TM} + c_{L,TM}) = \frac{1}{4}(c_{L,TE} + c_{H,TE}) = L_{+,TE}$$

$$L_{-,TM} = L_{H,TM}c_{H,TM} - L_{L,TM}c_{L,TM} == \frac{1}{4}(c_{H,TM} - c_{L,TM}) = \frac{1}{4}(c_{L,TE} - c_{H,TE}) = -L_{+,TE}$$

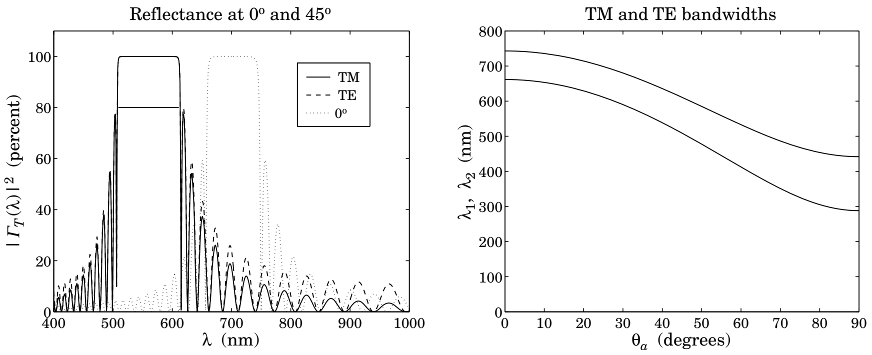


Fig. 8.13.2 Birefringent mirror with identical TM and TE reflection bands.

Because the computational algorithm (8.8.17) for the bandwidth does not depend on the sign of L_-, it follows that Eq. (8.8.17) will have the same solution for the TM and TE cases. The typical MATLAB code for this example was:

```
LH = 0.25; LL = 0.25;

na = [1.4; 1.4; 1.4];
nb = [1.4; 1.4; 1.4];
nH = [1.8; 1.8; 1.5];
nL = [1.5; 1.5; 1.8];

la0 = 700;
la = linspace(400,1000,601);

tha = 45;

N = 30;
n = [na, repmat([nH,nL], 1, N), nb];
L = [repmat([LH,LL], 1, N)];

Ge = 100*abs(multidiel(n, L, la/la0, tha, 'te')).^2;
Gm = 100*abs(multidiel(n, L, la/la0, tha, 'tm')).^2;
G0 = 100*abs(multidiel(n, L, la/la0)).^2;

plot(la,Gm,'-', la,Ge,'--', la,G0,':');
```

In Fig. 8.13.3, the low-index material is changed slightly to $\mathbf{n}_L = [1.5, 1.5, 1.9]$. The main behavior of the structure remains the same, except now the TM and TE bands are slightly different.

The MATLAB code used to compute the right graph was:

```
theta = linspace(0,90,361);                          % incident angles

F1e = []; F2e = [];
F1m = []; F2m = [];
```

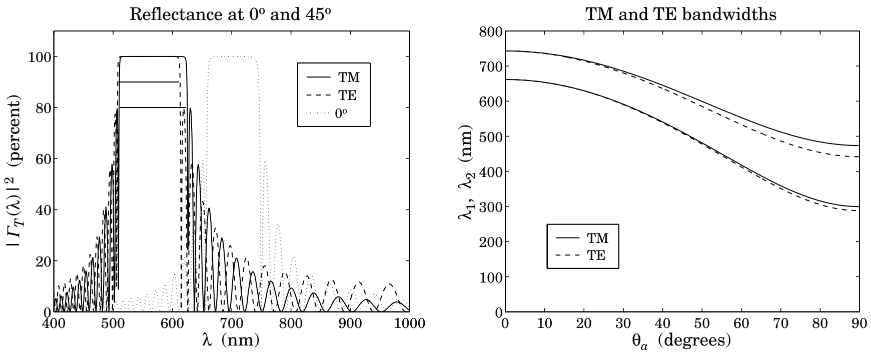


Fig. 8.13.3 Birefringent mirror with slightly different TM and TE reflection bands.

```
Ni = 3;                                                  % refinement iterations

for i=1:length(theta),
  [f1e,f2e] = omniband2(na,nH,nL,LH,LL,theta(i),'te',Ni);
  [f1m,f2m] = omniband2(na,nH,nL,LH,LL,theta(i),'tm',Ni);
  F1e = [F1e,f1e]; F2e = [F2e,f2e];
  F1m = [F1m,f1m]; F2m = [F2m,f2m];                      % frequency bandedges
end

la1e = la0 ./ F2e; la2e = la0 ./ F1e;                    % wavelength bandedges
la1m = la0 ./ F2m; la2m = la0 ./ F1m;

plot(theta,la1m,'-', theta,la2m,'-', theta,la1e,'--', theta,la2e,'--');
```

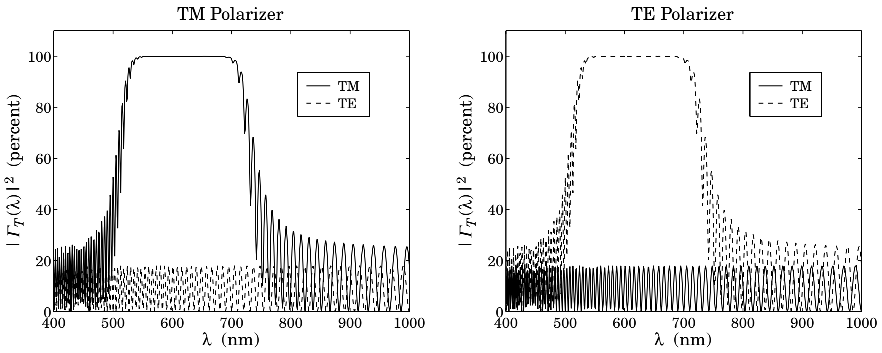
As the incident angle increases, not only does the TM band widen but it also becomes wider than the TE band—exactly the opposite behavior from the isotropic case. □

Example 8.13.3: *GBO Reflective Polarizer.* By choosing biaxial high/low layers whose refractive indices are mismatched only in the x or the y direction, one can design a mirror structure that reflects only the TM or only the TE polarization.

Fig. 8.13.4 shows the reflectance of an 80-bilayer mirror with $\mathbf{n}_H = [1.86, 1.57, 1.57]$ for the left graph, and $\mathbf{n}_H = [1.57, 1.86, 1.57]$ for the right one. In both graphs, the low index material is the same, with $\mathbf{n}_L = [1.57, 1.57, 1.57]$.

The angle of incidence was $\theta_a = 0^o$. The typical MATLAB code was:

```
LH = 0.25; LL = 0.25;

na = [1; 1; 1];
nb = [1; 1; 1];
nH = [1.86; 1.57; 1.57];
nL = [1.57; 1.57; 1.57];

la0 = 700;
la = linspace(400,1000,601);
```


TM Polarizer

TE Polarizer

Fig. 8.13.4 TM and TE mirror polarizers.

```
N = 80;
n = [na, repmat([nH,nL], 1, N), nb];
L = [repmat([LH,LL], 1, N)];
L = L .* linspace(1,0.75,2*N);                         % 25% thickness gradient

Ge = 100*abs(multidiel(n, L, la/la0, 0, 'te')).^2;
Gm = 100*abs(multidiel(n, L, la/la0, 0, 'tm')).^2;

plot(la,Gm,'-', la,Ge,'--');
```

A 25% thickness gradient was assumed in both cases. In the first case, the x-direction indices are different and the structure will act as a mirror for the TM polarization. The TE polarization will be reflected only by the air-high interface.

In the second case, the materials are matched in their y-direction indices and therefore, the structure becomes a mirror for the TE polarization, assuming as always that the plane of incidence is still the xz plane. □

Giant birefringent optics is a new paradigm in the design of multilayer mirrors and polarizers [658], offering increased flexibility in the control of reflected light. The recently manufactured *multilayer optical film* by 3M Corp. [1329] consists of hundreds to thousands of birefringent polymer layers with individual thicknesses of the order of a wavelength and total thickness of a sheet of paper. The optical working range of such films are between 400–2500 nm.

Applications include the design of efficient waveguides for transporting visible light over long distances and piping sunlight into interior rooms, reflective polarizers for improving liquid crystal displays, and other products, such as various optoelectronic components, cosmetics, and ”hot” and ”cold” mirrors for architectural and automotive windows.

8.14 Problems

8.1 Prove the reflectance and transmittance formulas (8.4.6) in FTIR.

8.2 *Computer Experiment—FTIR.* Reproduce the results and graphs of Figures 8.4.3–8.4.5.

8.3 *Computer Experiment—Surface Plasmon Resonance.* Reproduce the results and graphs of Figures 8.5.3–8.5.7.

8.4 Working with the electric and magnetic fields across an negative-index slab given by Eqs. (8.6.1) and (8.6.2), derive the reflection and transmission responses of the slab given in (8.6.8).

8.5 *Computer Experiment—Perfect Lens.* Study the sensitivity of the perfect lens property to the deviations from the ideal values of $\epsilon = -\epsilon_0$ and $\mu = -\mu_0$, and to the presence of losses by reproducing the results and graphs of Figures 8.6.3 and 8.6.4. You will need to implement the computational algorithm listed on page 326.

8.6 *Computer Experiment—Antireflection Coatings.* Reproduce the results and graphs of Figures 8.7.1–8.7.3.

8.7 *Computer Experiment—Omnidirectional Dielectric Mirrors.* Reproduce the results and graphs of Figures 8.8.2–8.8.10.

8.8 Derive the generalized Snel’s laws given in Eq. (8.10.10). Moreover, derive the Brewster angle expressions given in Eqs. (8.11.4) and (8.11.5).

8.9 *Computer Experiment—Brewster angles.* Study the variety of possible Brewster angles and reproduce the results and graphs of Example 8.11.1.

8.10 *Computer Experiment—Multilayer Birefringent Structures.* Reproduce the results and graphs of Figures 8.13.1–8.13.2.

9

Waveguides

Waveguides are used to transfer electromagnetic power efficiently from one point in space to another. Some common guiding structures are shown in the figure below. These include the typical coaxial cable, the two-wire and mictrostrip transmission lines, hollow conducting waveguides, and optical fibers.

In practice, the choice of structure is dictated by: (a) the desired operating frequency band, (b) the amount of power to be transferred, and (c) the amount of transmission losses that can be tolerated.

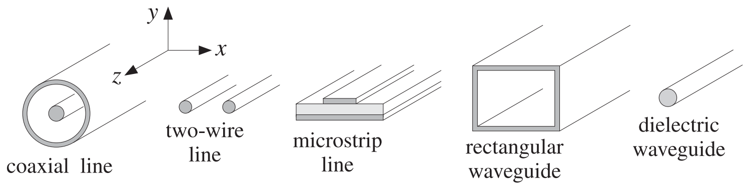


Fig. 9.0.1 Typical waveguiding structures.

Coaxial cables are widely used to connect RF components. Their operation is practical for frequencies below 3 GHz. Above that the losses are too excessive. For example, the attenuation might be 3 dB per 100 m at 100 MHz, but 10 dB/100 m at 1 GHz, and 50 dB/100 m at 10 GHz. Their power rating is typically of the order of one kilowatt at 100 MHz, but only 200 W at 2 GHz, being limited primarily because of the heating of the coaxial conductors and of the dielectric between the conductors (dielectric voltage breakdown is usually a secondary factor.) However, special short-length coaxial cables do exist that operate in the 40 GHz range.

Another issue is the single-mode operation of the line. At higher frequencies, in order to prevent higher modes from being launched, the diameters of the coaxial conductors must be reduced, diminishing the amount of power that can be transmitted.

Two-wire lines are not used at microwave frequencies because they are not shielded and can radiate. One typical use is for connecting indoor antennas to TV sets. Microstrip lines are used widely in microwave integrated circuits.

Rectangular waveguides are used routinely to transfer large amounts of microwave power at frequencies greater than 3 GHz. For example at 5 GHz, the transmitted power might be one megawatt and the attenuation only 4 dB/100 m.

Optical fibers operate at optical and infrared frequencies, allowing a very wide bandwidth. Their losses are very low, typically, 0.2 dB/km. The transmitted power is of the order of milliwatts.

9.1 Longitudinal-Transverse Decompositions

In a waveguiding system, we are looking for solutions of Maxwell's equations that are propagating along the guiding direction (the z direction) and are confined in the near vicinity of the guiding structure. Thus, the electric and magnetic fields are assumed to have the form:

$$\boxed{\begin{aligned} \mathbf{E}(x,y,z,t) &= \mathbf{E}(x,y)\,e^{j\omega t-j\beta z} \\ \mathbf{H}(x,y,z,t) &= \mathbf{H}(x,y)\,e^{j\omega t-j\beta z} \end{aligned}} \tag{9.1.1}$$

where β is the propagation wavenumber along the guide direction. The corresponding wavelength, called the *guide wavelength*, is denoted by $\lambda_g = 2\pi/\beta$.

The precise relationship between ω and β depends on the type of waveguiding structure and the particular propagating mode. Because the fields are confined in the transverse directions (the x,y directions,) they cannot be uniform (except in very simple structures) and will have a non-trivial dependence on the transverse coordinates x and y. Next, we derive the equations for the phasor amplitudes $\mathbf{E}(x,y)$ and $\mathbf{H}(x,y)$.

Because of the preferential role played by the guiding direction z, it proves convenient to decompose Maxwell's equations into components that are *longitudinal*, that is, along the z-direction, and components that are *transverse*, along the x,y directions. Thus, we decompose:

$$\mathbf{E}(x,y) = \underbrace{\hat{\mathbf{x}}\,E_x(x,y)+\hat{\mathbf{y}}\,E_y(x,y)}_{\text{transverse}} + \underbrace{\hat{\mathbf{z}}\,E_z(x,y)}_{\text{longitudinal}} \equiv \mathbf{E}_T(x,y)+\hat{\mathbf{z}}\,E_z(x,y) \tag{9.1.2}$$

In a similar fashion we may decompose the gradient operator:

$$\nabla = \underbrace{\hat{\mathbf{x}}\,\partial_x + \hat{\mathbf{y}}\,\partial_y}_{\text{transverse}} + \hat{\mathbf{z}}\,\partial_z = \nabla_T + \hat{\mathbf{z}}\,\partial_z = \nabla_T - j\beta\,\hat{\mathbf{z}} \tag{9.1.3}$$

where we made the replacement $\partial_z \to -j\beta$ because of the assumed z-dependence. Introducing these decompositions into the source-free Maxwell's equations we have:

$$\begin{aligned}
\nabla\times\mathbf{E} &= -j\omega\mu\mathbf{H} & \qquad & (\nabla_T - j\beta\hat{\mathbf{z}})\times(\mathbf{E}_T+\hat{\mathbf{z}}\,E_z) = -j\omega\mu\,(\mathbf{H}_T+\hat{\mathbf{z}}\,H_z) \\
\nabla\times\mathbf{H} &= j\omega\epsilon\mathbf{E} & & (\nabla_T - j\beta\hat{\mathbf{z}})\times(\mathbf{H}_T+\hat{\mathbf{z}}\,H_z) = j\omega\epsilon\,(\mathbf{E}_T+\hat{\mathbf{z}}\,E_z) \\
\nabla\cdot\mathbf{E} &= 0 & \Rightarrow & (\nabla_T - j\beta\hat{\mathbf{z}})\cdot(\mathbf{E}_T+\hat{\mathbf{z}}\,E_z) = 0 \\
\nabla\cdot\mathbf{H} &= 0 & & (\nabla_T - j\beta\hat{\mathbf{z}})\cdot(\mathbf{H}_T+\hat{\mathbf{z}}\,H_z) = 0
\end{aligned} \tag{9.1.4}$$

where ϵ,μ denote the permittivities of the medium in which the fields propagate, for example, the medium between the coaxial conductors in a coaxial cable, or the medium within the hollow rectangular waveguide. This medium is assumed to be lossless for now.

We note that $\hat{\mathbf{z}}\cdot\hat{\mathbf{z}} = 1$, $\hat{\mathbf{z}}\times\hat{\mathbf{z}} = 0$, $\hat{\mathbf{z}}\cdot\mathbf{E}_T = 0$, $\hat{\mathbf{z}}\cdot\nabla_T E_z = 0$ and that $\hat{\mathbf{z}}\times\mathbf{E}_T$ and $\hat{\mathbf{z}}\times\nabla_T E_z$ are transverse while $\nabla_T\times\mathbf{E}_T$ is longitudinal. Indeed, we have:

$$\begin{aligned}
\hat{\mathbf{z}}\times\mathbf{E}_T &= \hat{\mathbf{z}}\times(\hat{\mathbf{x}}\,E_x+\hat{\mathbf{y}}\,E_y) = \hat{\mathbf{y}}\,E_x - \hat{\mathbf{x}}\,E_y \\
\nabla_T\times\mathbf{E}_T &= (\hat{\mathbf{x}}\,\partial_x+\hat{\mathbf{y}}\,\partial_y)\times(\hat{\mathbf{x}}\,E_x+\hat{\mathbf{y}}\,E_y) = \hat{\mathbf{z}}\,(\partial_x E_y - \partial_y E_x)
\end{aligned}$$

Using these properties and equating longitudinal and transverse parts in the two sides of Eq. (9.1.4), we obtain the equivalent set of Maxwell equations:

$$\begin{aligned}
\nabla_T E_z\times\hat{\mathbf{z}} - j\beta\,\hat{\mathbf{z}}\times\mathbf{E}_T &= -j\omega\mu\mathbf{H}_T \\
\nabla_T H_z\times\hat{\mathbf{z}} - j\beta\,\hat{\mathbf{z}}\times\mathbf{H}_T &= j\omega\epsilon\mathbf{E}_T \\
\nabla_T\times\mathbf{E}_T + j\omega\mu\,\hat{\mathbf{z}}\,H_z &= 0 \\
\nabla_T\times\mathbf{H}_T - j\omega\epsilon\,\hat{\mathbf{z}}\,E_z &= 0 \\
\nabla_T\cdot\mathbf{E}_T - j\beta E_z &= 0 \\
\nabla_T\cdot\mathbf{H}_T - j\beta H_z &= 0
\end{aligned} \tag{9.1.5}$$

Depending on whether both, one, or none of the longitudinal components are zero, we may classify the solutions as transverse electric and magnetic (TEM), transverse electric (TE), transverse magnetic (TM), or hybrid:

$$\boxed{\begin{array}{ll}
E_z = 0,\ H_z = 0, & \text{TEM modes} \\
E_z = 0,\ H_z \neq 0, & \text{TE or H modes} \\
E_z \neq 0,\ H_z = 0, & \text{TM or E modes} \\
E_z \neq 0,\ H_z \neq 0, & \text{hybrid or HE or EH modes}
\end{array}}$$

In the case of TEM modes, which are the dominant modes in two-conductor transmission lines such as the coaxial cable, the fields are purely transverse and the solution of Eq. (9.1.5) reduces to an equivalent two-dimensional electrostatic problem. We will discuss this case later on.

In all other cases, at least one of the longitudinal fields E_z, H_z is non-zero. It is then possible to express the transverse field components $\mathbf{E}_T$, $\mathbf{H}_T$ in terms of the longitudinal ones, E_z, H_z.

Forming the cross-product of the second of equations (9.1.5) with $\hat{\mathbf{z}}$ and using the BAC-CAB vector identity, $\hat{\mathbf{z}}\times(\hat{\mathbf{z}}\times\mathbf{H}_T) = \hat{\mathbf{z}}(\hat{\mathbf{z}}\cdot\mathbf{H}_T) - \mathbf{H}_T(\hat{\mathbf{z}}\cdot\hat{\mathbf{z}}) = -\mathbf{H}_T$, and similarly, $\hat{\mathbf{z}}\times(\nabla_T H_z\times\hat{\mathbf{z}}) = \nabla_T H_z$, we obtain:

$$\nabla_T H_z + j\beta\mathbf{H}_T = j\omega\epsilon\,\hat{\mathbf{z}}\times\mathbf{E}_T$$

Thus, the first two of (9.1.5) may be thought of as a linear system of two equations in the two unknowns $\hat{\mathbf{z}}\times\mathbf{E}_T$ and $\mathbf{H}_T$, that is,

$$\begin{aligned}
\beta\,\hat{\mathbf{z}}\times\mathbf{E}_T - \omega\mu\mathbf{H}_T &= j\hat{\mathbf{z}}\times\nabla_T E_z \\
\omega\epsilon\,\hat{\mathbf{z}}\times\mathbf{E}_T - \beta\mathbf{H}_T &= -j\nabla_T H_z
\end{aligned} \tag{9.1.6}$$

The solution of this system is:

$$\begin{aligned}\hat{\mathbf{z}}\times\boldsymbol{E}_T &= -\frac{j\beta}{k_c^2}\,\hat{\mathbf{z}}\times\nabla_T E_z - \frac{j\omega\mu}{k_c^2}\,\nabla_T H_z\\ \boldsymbol{H}_T &= -\frac{j\omega\epsilon}{k_c^2}\,\hat{\mathbf{z}}\times\nabla_T E_z - \frac{j\beta}{k_c^2}\,\nabla_T H_z\end{aligned}\tag{9.1.7}$$

where we defined the so-called *cutoff wavenumber* k_c by:

$$\boxed{k_c^2 = \omega^2\epsilon\mu - \beta^2 = \frac{\omega^2}{c^2} - \beta^2 = k^2 - \beta^2}\quad\text{(cutoff wavenumber)}\tag{9.1.8}$$

The quantity $k = \omega/c = \omega\sqrt{\epsilon\mu}$ is the wavenumber a uniform plane wave would have in the propagation medium ϵ,μ.

Although k_c^2 stands for the difference $\omega^2\epsilon\mu - \beta^2$, it turns out that the boundary conditions for each waveguide type force k_c^2 to take on certain values, which can be positive, negative, or zero, and characterize the propagating modes. For example, in a dielectric waveguide k_c^2 is positive inside the guide and negative outside it; in a hollow conducting waveguide k_c^2 takes on certain quantized positive values; in a TEM line, k_c^2 is zero. Some related definitions are the *cutoff frequency* and the *cutoff wavelength* defined as follows:

$$\boxed{\omega_c = ck_c\,,\quad \lambda_c = \frac{2\pi}{k_c}}\quad\text{(cutoff frequency and wavelength)}\tag{9.1.9}$$

We can then express β in terms of ω and ω_c, or ω in terms of β and ω_c. Taking the positive square roots of Eq. (9.1.8), we have:

$$\boxed{\beta = \frac{1}{c}\sqrt{\omega^2-\omega_c^2} = \frac{\omega}{c}\sqrt{1-\frac{\omega_c^2}{\omega^2}}\quad\text{and}\quad \omega = \sqrt{\omega_c^2+\beta^2c^2}}\tag{9.1.10}$$

Often, Eq. (9.1.10) is expressed in terms of the wavelengths $\lambda = 2\pi/k = 2\pi c/\omega$, $\lambda_c = 2\pi/k_c$, and $\lambda_g = 2\pi/\beta$. It follows from $k^2 = k_c^2 + \beta^2$ that

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_c^2} + \frac{1}{\lambda_g^2}\quad\Rightarrow\quad \lambda_g = \frac{\lambda}{\sqrt{1-\dfrac{\lambda^2}{\lambda_c^2}}}\tag{9.1.11}$$

Note that λ is related to the free-space wavelength $\lambda_0 = 2\pi c_0/\omega = c_0/f$ by the refractive index of the dielectric material $\lambda = \lambda_0/n$.

It is convenient at this point to introduce the *transverse impedances* for the TE and TM modes by the definitions:

$$\boxed{\eta_{TE} = \frac{\omega\mu}{\beta} = \eta\,\frac{\omega}{\beta c}\,,\quad \eta_{TM} = \frac{\beta}{\omega\epsilon} = \eta\,\frac{\beta c}{\omega}}\quad\text{(TE and TM impedances)}\tag{9.1.12}$$

where the medium impedance is $\eta = \sqrt{\mu/\epsilon}$, so that $\eta/c = \mu$ and $\eta c = 1/\epsilon$. We note the properties:

$$\boxed{\eta_{TE}\eta_{TM} = \eta^2\,,\quad \frac{\eta_{TE}}{\eta_{TM}} = \frac{\omega^2}{\beta^2c^2}}\tag{9.1.13}$$

Because $\beta c/\omega = \sqrt{1-\omega_c^2/\omega^2}$, we can write also:

$$\boxed{\eta_{TE} = \frac{\eta}{\sqrt{1-\dfrac{\omega_c^2}{\omega^2}}}\,,\quad \eta_{TM} = \eta\sqrt{1-\frac{\omega_c^2}{\omega^2}}}\tag{9.1.14}$$

With these definitions, we may rewrite Eq. (9.1.7) as follows:

$$\begin{aligned}\hat{\mathbf{z}}\times\boldsymbol{E}_T &= -\frac{j\beta}{k_c^2}\left(\hat{\mathbf{z}}\times\nabla_T E_z + \eta_{TE}\nabla_T H_z\right)\\ \boldsymbol{H}_T &= -\frac{j\beta}{k_c^2}\left(\frac{1}{\eta_{TM}}\,\hat{\mathbf{z}}\times\nabla_T E_z + \nabla_T H_z\right)\end{aligned}\tag{9.1.15}$$

Using the result $\hat{\mathbf{z}}\times(\hat{\mathbf{z}}\times\boldsymbol{E}_T) = -\boldsymbol{E}_T$, we solve for $\boldsymbol{E}_T$ and $\boldsymbol{H}_T$:

$$\boxed{\begin{aligned}\boldsymbol{E}_T &= -\frac{j\beta}{k_c^2}\left(\nabla_T E_z - \eta_{TE}\,\hat{\mathbf{z}}\times\nabla_T H_z\right)\\ \boldsymbol{H}_T &= -\frac{j\beta}{k_c^2}\left(\nabla_T H_z + \frac{1}{\eta_{TM}}\,\hat{\mathbf{z}}\times\nabla_T E_z\right)\end{aligned}}\quad\text{(transverse fields)}\tag{9.1.16}$$

An alternative and useful way of writing these equations is to form the following linear combinations, which are equivalent to Eq. (9.1.6):

$$\boxed{\begin{aligned}\boldsymbol{H}_T - \frac{1}{\eta_{TM}}\,\hat{\mathbf{z}}\times\boldsymbol{E}_T &= \frac{j}{\beta}\,\nabla_T H_z\\ \boldsymbol{E}_T - \eta_{TE}\,\boldsymbol{H}_T\times\hat{\mathbf{z}} &= \frac{j}{\beta}\,\nabla_T E_z\end{aligned}}\tag{9.1.17}$$

So far we only used the first two of Maxwell's equations (9.1.5) and expressed $\boldsymbol{E}_T, \boldsymbol{H}_T$ in terms of E_z, H_z. Using (9.1.16), it is easily shown that the left-hand sides of the remaining four of Eqs. (9.1.5) take the forms:

$$\begin{aligned}\nabla_T\times\boldsymbol{E}_T + j\omega\mu\,\hat{\mathbf{z}}\,H_z &= \frac{j\omega\mu}{k_c^2}\,\hat{\mathbf{z}}\,(\nabla_T^2 H_z + k_c^2 H_z)\\ \nabla_T\times\boldsymbol{H}_T - j\omega\epsilon\,\hat{\mathbf{z}}\,E_z &= -\frac{j\omega\epsilon}{k_c^2}\,\hat{\mathbf{z}}\,(\nabla_T^2 E_z + k_c^2 E_z)\\ \nabla_T\cdot\boldsymbol{E}_T - j\beta E_z &= -\frac{j\beta}{k_c^2}\,(\nabla_T^2 E_z + k_c^2 E_z)\\ \nabla_T\cdot\boldsymbol{H}_T - j\beta H_z &= -\frac{j\beta}{k_c^2}\,(\nabla_T^2 H_z + k_c^2 H_z)\end{aligned}$$

where ∇_T^2 is the two-dimensional Laplacian operator:

$$\nabla_T^2 = \boldsymbol{\nabla}_T \cdot \boldsymbol{\nabla}_T = \partial_x^2 + \partial_y^2 \tag{9.1.18}$$

and we used the vectorial identities $\boldsymbol{\nabla}_T \times \boldsymbol{\nabla}_T E_z = 0$, $\boldsymbol{\nabla}_T \times (\hat{\mathbf{z}} \times \boldsymbol{\nabla}_T H_z) = \hat{\mathbf{z}}\,\nabla_T^2 H_z$, and $\boldsymbol{\nabla}_T \cdot (\hat{\mathbf{z}} \times \boldsymbol{\nabla}_T H_z) = 0$.

It follows that in order to satisfy all of the last four of Maxwell's equations (9.1.5), it is necessary that the longitudinal fields $E_z(x,y), H_z(x,y)$ satisfy the two-dimensional Helmholtz equations:

$$\boxed{\begin{aligned} &\nabla_T^2 E_z + k_c^2 E_z = 0 \\ &\nabla_T^2 H_z + k_c^2 H_z = 0 \end{aligned}} \qquad \text{(Helmholtz equations)} \tag{9.1.19}$$

These equations are to be solved subject to the appropriate *boundary conditions* for each waveguide type. Once, the fields E_z, H_z are known, the transverse fields $\boldsymbol{E}_T, \boldsymbol{H}_T$ are computed from Eq. (9.1.16), resulting in a complete solution of Maxwell's equations for the guiding structure. To get the full x, y, z, t dependence of the propagating fields, the above solutions must be multiplied by the factor $e^{j\omega t - j\beta z}$.

The cross-sections of practical waveguiding systems have either cartesian or cylindrical symmetry, such as the rectangular waveguide or the coaxial cable. Below, we summarize the form of the above solutions in the two types of coordinate systems.

Cartesian Coordinates

The cartesian component version of Eqs. (9.1.16) and (9.1.19) is straightforward. Using the identity $\hat{\mathbf{z}} \times \boldsymbol{\nabla}_T H_z = \hat{\mathbf{y}}\,\partial_x H_z - \hat{\mathbf{x}}\,\partial_y H_z$, we obtain for the longitudinal components:

$$\boxed{\begin{aligned} &(\partial_x^2 + \partial_y^2) E_z + k_c^2 E_z = 0 \\ &(\partial_x^2 + \partial_y^2) H_z + k_c^2 H_z = 0 \end{aligned}} \tag{9.1.20}$$

Eq. (9.1.16) becomes for the transverse components:

$$\boxed{\begin{aligned} E_x &= -\frac{j\beta}{k_c^2}(\partial_x E_z + \eta_{TE}\,\partial_y H_z) \\ E_y &= -\frac{j\beta}{k_c^2}(\partial_y E_z - \eta_{TE}\,\partial_x H_z) \end{aligned}}\,, \quad \boxed{\begin{aligned} H_x &= -\frac{j\beta}{k_c^2}(\partial_x H_z - \frac{1}{\eta_{TM}}\,\partial_y E_z) \\ H_y &= -\frac{j\beta}{k_c^2}(\partial_y H_z + \frac{1}{\eta_{TM}}\,\partial_x E_z) \end{aligned}} \tag{9.1.21}$$

Cylindrical Coordinates

The relationship between cartesian and cylindrical coordinates is shown in Fig. 9.1.1. From the triangle in the figure, we have $x = \rho\cos\phi$ and $y = \rho\sin\phi$. The transverse gradient and Laplace operator are in cylindrical coordinates:

$$\boldsymbol{\nabla}_T = \hat{\boldsymbol{\rho}}\,\frac{\partial}{\partial\rho} + \hat{\boldsymbol{\phi}}\,\frac{1}{\rho}\frac{\partial}{\partial\phi}\,, \qquad \nabla_T^2 = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} \tag{9.1.22}$$

The Helmholtz equations (9.1.19) now read:

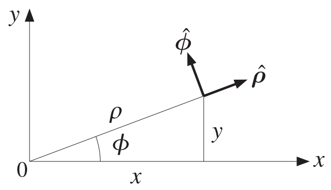


Fig. 9.1.1 Cylindrical coordinates.

$$\boxed{\begin{aligned} &\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial E_z}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 E_z}{\partial\phi^2} + k_c^2 E_z = 0 \\ &\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial H_z}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 H_z}{\partial\phi^2} + k_c^2 H_z = 0 \end{aligned}} \tag{9.1.23}$$

Noting that $\hat{\mathbf{z}} \times \hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\phi}}$ and $\hat{\mathbf{z}} \times \hat{\boldsymbol{\phi}} = -\hat{\boldsymbol{\rho}}$, we obtain:

$$\hat{\mathbf{z}} \times \boldsymbol{\nabla}_T H_z = \hat{\boldsymbol{\phi}}\,(\partial_\rho H_z) - \hat{\boldsymbol{\rho}}\,\frac{1}{\rho}\,(\partial_\phi H_z)$$

The decomposition of a transverse vector is $\boldsymbol{E}_T = \hat{\boldsymbol{\rho}}\,E_\rho + \hat{\boldsymbol{\phi}}\,E_\phi$. The cylindrical coordinates version of (9.1.16) are:

$$\boxed{\begin{aligned} E_\rho &= -\frac{j\beta}{k_c^2}(\partial_\rho E_z - \eta_{TE}\frac{1}{\rho}\partial_\phi H_z) \\ E_\phi &= -\frac{j\beta}{k_c^2}(\frac{1}{\rho}\partial_\phi E_z + \eta_{TE}\partial_\rho H_z) \end{aligned}}\,, \quad \boxed{\begin{aligned} H_\rho &= -\frac{j\beta}{k_c^2}(\partial_\rho H_z + \frac{1}{\eta_{TM}\rho}\partial_\phi E_z) \\ H_\phi &= -\frac{j\beta}{k_c^2}(\frac{1}{\rho}\partial_\phi H_z - \frac{1}{\eta_{TM}}\partial_\rho E_z) \end{aligned}} \tag{9.1.24}$$

For either coordinate system, the equations for $\boldsymbol{H}_T$ may be obtained from those of $\boldsymbol{E}_T$ by a so-called *duality transformation*, that is, making the substitutions:

$$\boldsymbol{E} \to \boldsymbol{H}, \quad \boldsymbol{H} \to -\boldsymbol{E}, \quad \epsilon \to \mu, \quad \mu \to \epsilon \qquad \text{(duality transformation)} \tag{9.1.25}$$

These imply that $\eta \to \eta^{-1}$ and $\eta_{TE} \to \eta_{TM}^{-1}$. Duality is discussed in greater detail in Sec. 17.2.

9.2 Power Transfer and Attenuation

With the field solutions at hand, one can determine the amount of power transmitted along the guide, as well as the transmission losses. The total power carried by the fields along the guide direction is obtained by integrating the z-component of the Poynting vector over the cross-sectional area of the guide:

$$\boxed{P_T = \int_S \mathcal{P}_z \, dS}\,, \quad \text{where} \quad \mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*)\cdot\hat{\mathbf{z}} \tag{9.2.1}$$

It is easily verified that only the transverse components of the fields contribute to the power flow, that is, $\mathcal{P}_z$ can be written in the form:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\mathbf{E}_T\times\mathbf{H}_T^*)\cdot\hat{\mathbf{z}} \tag{9.2.2}$$

For waveguides with conducting walls, the transmission losses are due primarily to ohmic losses in (a) the conductors and (b) the dielectric medium filling the space between the conductors and in which the fields propagate. In dielectric waveguides, the losses are due to absorption and scattering by imperfections.

The transmission losses can be quantified by replacing the propagation wavenumber β by its complex-valued version $\beta_c = \beta - j\alpha$, where α is the attenuation constant. The z-dependence of all the field components is replaced by:

$$e^{-j\beta z} \quad\rightarrow\quad e^{-j\beta_c z} = e^{-(\alpha+j\beta)z} = e^{-\alpha z}e^{-j\beta z} \tag{9.2.3}$$

The quantity α is the sum of the attenuation constants arising from the various loss mechanisms. For example, if α_d and α_c are the attenuations due to the ohmic losses in the dielectric and in the conducting walls, then

$$\alpha = \alpha_d + \alpha_c \tag{9.2.4}$$

The ohmic losses in the dielectric can be characterized either by its loss tangent, say $\tan\delta$, or by its conductivity σ_d—the two being related by $\sigma_d = \omega\epsilon\tan\delta$. The effective dielectric constant of the medium is then $\epsilon(\omega) = \epsilon - j\sigma_d/\omega = \epsilon(1 - j\tan\delta)$. The corresponding complex-valued wavenumber β_c is obtained by the replacement:

$$\beta = \sqrt{\omega^2\mu\epsilon - k_c^2} \quad\rightarrow\quad \beta_c = \sqrt{\omega^2\mu\epsilon(\omega) - k_c^2}$$

For weakly conducting dielectrics, we may make the approximation:

$$\beta_c = \sqrt{\omega^2\mu\epsilon\Big(1 - j\frac{\sigma_d}{\omega\epsilon}\Big) - k_c^2} = \sqrt{\beta^2 - j\omega\mu\sigma_d} = \beta\sqrt{1 - j\frac{\omega\mu\sigma_d}{\beta^2}} \simeq \beta - j\frac{1}{2}\,\sigma_d\frac{\omega\mu}{\beta}$$

Recalling the definition $\eta_{TE} = \omega\mu/\beta$, we obtain for the attenuation constant:

$$\boxed{\alpha_d = \frac{1}{2}\,\sigma_d\eta_{TE} = \frac{1}{2}\,\frac{\omega^2}{\beta c^2}\tan\delta = \frac{\omega\tan\delta}{2c\sqrt{1-\omega_c^2/\omega^2}}} \qquad \text{(dielectric losses)} \tag{9.2.5}$$

which is similar to Eq. (2.7.2), but with the replacement $\eta_d \rightarrow \eta_{TE}$.

The conductor losses are more complicated to calculate. In practice, the following approximate procedure is adequate. First, the fields are determined on the assumption that the conductors are perfect.

Second, the magnetic fields on the conductor surfaces are determined and the corresponding induced surface currents are calculated by $\mathbf{J}_s = \hat{\mathbf{n}}\times\mathbf{H}$, where $\hat{\mathbf{n}}$ is the outward normal to the conductor.

Third, the ohmic losses per unit conductor area are calculated by Eq. (2.8.7). Figure 9.2.1 shows such an infinitesimal conductor area $dA = dl\,dz$, where dl is along the cross-sectional periphery of the conductor. Applying Eq. (2.8.7) to this area, we have:

$$\frac{dP_{\text{loss}}}{dA} = \frac{dP_{\text{loss}}}{dldz} = \frac{1}{2}R_s|\mathbf{J}_s|^2 \tag{9.2.6}$$

where R_s is the surface resistance of the conductor given by Eq. (2.8.4),

$$R_s = \sqrt{\frac{\omega\mu}{2\sigma}} = \eta\sqrt{\frac{\omega\epsilon}{2\sigma}} = \frac{1}{2}\delta\omega\mu\,, \qquad \delta = \sqrt{\frac{2}{\omega\mu\sigma}} = \text{skin depth} \tag{9.2.7}$$

Integrating Eq. (9.2.6) around the periphery of the conductor gives the power loss per unit z-length due to that conductor. Adding similar terms for all the other conductors gives the total power loss per unit z-length:

$$\boxed{P'_{\text{loss}} = \frac{dP_{\text{loss}}}{dz} = \oint_{C_a}\frac{1}{2}\,R_s|\mathbf{J}_s|^2\,dl + \oint_{C_b}\frac{1}{2}\,R_s|\mathbf{J}_s|^2\,dl} \tag{9.2.8}$$

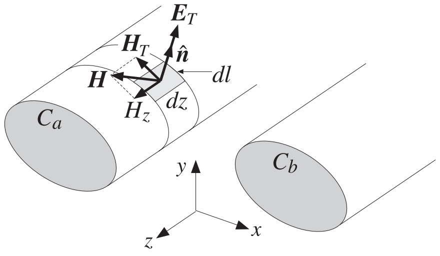


Fig. 9.2.1 Conductor surface absorbs power from the propagating fields.

where C_a and C_b indicate the peripheries of the conductors. Finally, the corresponding attenuation coefficient is calculated from Eq. (2.6.22):

$$\boxed{\alpha_c = \frac{P'_{\text{loss}}}{2P_T}} \qquad \text{(conductor losses)} \tag{9.2.9}$$

Equations (9.2.1)–(9.2.9) provide a systematic methodology by which to calculate the transmitted power and attenuation losses in waveguides. We will apply it to several examples later on.

9.3 TEM, TE, and TM modes

The general solution described by Eqs. (9.1.16) and (9.1.19) is a hybrid solution with nonzero E_z and H_z components. Here, we look at the specialized forms of these equations in the cases of TEM, TE, and TM modes.

One common property of all three types of modes is that the transverse fields $\boldsymbol{E}_T, \boldsymbol{H}_T$ are related to each other in the same way as in the case of uniform plane waves propagating in the z-direction, that is, they are perpendicular to each other, their cross-product points in the z-direction, and they satisfy:

$$\boxed{\boldsymbol{H}_T = \frac{1}{\eta_T}\,\hat{\mathbf{z}}\times\boldsymbol{E}_T} \tag{9.3.1}$$

where η_T is the transverse impedance of the particular mode type, that is, $\eta, \eta_{TE}, \eta_{TM}$ in the TEM, TE, and TM cases.

Because of Eq. (9.3.1), the power flow per unit cross-sectional area described by the Poynting vector $\mathcal{P}_z$ of Eq. (9.2.2) takes the simple form in all three cases:

$$\boxed{\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\boldsymbol{E}_T\times\boldsymbol{H}_T^*)\cdot\hat{\mathbf{z}} = \frac{1}{2\eta_T}|\boldsymbol{E}_T|^2 = \frac{1}{2}\,\eta_T|\boldsymbol{H}_T|^2} \tag{9.3.2}$$

TEM modes

In TEM modes, both E_z and H_z vanish, and the fields are fully transverse. One can set $E_z = H_z = 0$ in Maxwell equations (9.1.5), or equivalently in (9.1.16), or in (9.1.17).

From any point view, one obtains the condition $k_c^2 = 0$, or $\omega = \beta c$. For example, if the right-hand sides of Eq. (9.1.17) vanish, the consistency of the system requires that $\eta_{TE} = \eta_{TM}$, which by virtue of Eq. (9.1.13) implies $\omega = \beta c$. It also implies that η_{TE}, η_{TM} must both be equal to the medium impedance η. Thus, the electric and magnetic fields satisfy:

$$\boxed{\boldsymbol{H}_T = \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}_T} \tag{9.3.3}$$

These are the same as in the case of a uniform plane wave, except here the fields are not uniform and may have a non-trivial x, y dependence. The electric field $\boldsymbol{E}_T$ is determined from the rest of Maxwell's equations (9.1.5), which read:

$$\boxed{\begin{aligned} &\boldsymbol{\nabla}_T\times\boldsymbol{E}_T = 0 \\ &\boldsymbol{\nabla}_T\cdot\boldsymbol{E}_T = 0 \end{aligned}} \tag{9.3.4}$$

These are recognized as the field equations of an equivalent two-dimensional *electrostatic* problem. Once this electrostatic solution is found, $\boldsymbol{E}_T(x,y)$, the magnetic field is constructed from Eq. (9.3.3). The time-varying propagating fields will be given by Eq. (9.1.1), with $\omega = \beta c$. (For backward moving fields, replace β by $-\beta$.)

We explore this electrostatic point of view further in Sec. 10.1 and discuss the cases of the coaxial, two-wire, and strip lines. Because of the relationship between $\boldsymbol{E}_T$ and $\boldsymbol{H}_T$, the Poynting vector $\mathcal{P}_z$ of Eq. (9.2.2) will be:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\boldsymbol{E}_T\times\boldsymbol{H}_T^*)\cdot\hat{\mathbf{z}} = \frac{1}{2\eta}|\boldsymbol{E}_T|^2 = \frac{1}{2}\,\eta|\boldsymbol{H}_T|^2 \tag{9.3.5}$$

TE modes

TE modes are characterized by the conditions $E_z = 0$ and $H_z \neq 0$. It follows from the second of Eqs. (9.1.17) that $\boldsymbol{E}_T$ is completely determined from $\boldsymbol{H}_T$, that is, $\boldsymbol{E}_T = \eta_{TE}\boldsymbol{H}_T\times\hat{\mathbf{z}}$.

The field $\boldsymbol{H}_T$ is determined from the second of (9.1.16). Thus, all field components for TE modes are obtained from the equations:

$$\boxed{\begin{aligned} &\nabla_T^2 H_z + k_c^2 H_z = 0 \\ &\boldsymbol{H}_T = -\frac{j\beta}{k_c^2}\,\boldsymbol{\nabla}_T H_z \\ &\boldsymbol{E}_T = \eta_{TE}\,\boldsymbol{H}_T\times\hat{\mathbf{z}} \end{aligned}} \qquad \text{(TE modes)} \tag{9.3.6}$$

The relationship of $\boldsymbol{E}_T$ and $\boldsymbol{H}_T$ is identical to that of uniform plane waves propagating in the z-direction, except the wave impedance is replaced by η_{TE}. The Poynting vector of Eq. (9.2.2) then takes the form:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\boldsymbol{E}_T\times\boldsymbol{H}_T^*)\cdot\hat{\mathbf{z}} = \frac{1}{2\eta_{TE}}|\boldsymbol{E}_T|^2 = \frac{1}{2}\,\eta_{TE}|\boldsymbol{H}_T|^2 = \frac{1}{2}\,\eta_{TE}\,\frac{\beta^2}{k_c^4}\,|\boldsymbol{\nabla}_T H_z|^2 \tag{9.3.7}$$

The cartesian coordinate version of Eq. (9.3.6) is:

$$\boxed{\begin{aligned} &(\partial_x^2 + \partial_y^2)H_z + k_c^2 H_z = 0 \\ &H_x = -\frac{j\beta}{k_c^2}\,\partial_x H_z\,, \quad H_y = -\frac{j\beta}{k_c^2}\,\partial_y H_z \\ &E_x = \eta_{TE}H_y\,, \quad E_y = -\eta_{TE}H_x \end{aligned}} \tag{9.3.8}$$

And, the cylindrical coordinate version:

$$\boxed{\begin{aligned} &\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial H_z}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 H_z}{\partial\phi^2} + k_c^2 H_z = 0 \\ &H_\rho = -\frac{j\beta}{k_c^2}\frac{\partial H_z}{\partial\rho}\,, \quad H_\phi = -\frac{j\beta}{k_c^2}\frac{1}{\rho}\frac{\partial H_z}{\partial\phi} \\ &E_\rho = \eta_{TE}H_\phi\,, \quad E_\phi = -\eta_{TE}H_\rho \end{aligned}} \tag{9.3.9}$$

where we used $\boldsymbol{H}_T\times\hat{\mathbf{z}} = (\hat{\boldsymbol{\rho}}H_\rho + \hat{\boldsymbol{\phi}}H_\phi)\times\hat{\mathbf{z}} = -\hat{\boldsymbol{\phi}}H_\rho + \hat{\boldsymbol{\rho}}H_\phi$.

TM modes

TM modes have $H_z = 0$ and $E_z \neq 0$. It follows from the first of Eqs. (9.1.17) that $\boldsymbol{H}_T$ is completely determined from $\boldsymbol{E}_T$, that is, $\boldsymbol{H}_T = \eta_{TM}^{-1}\hat{\mathbf{z}}\times\boldsymbol{E}_T$. The field $\boldsymbol{E}_T$ is determined from the first of (9.1.16), so that all field components for TM modes are obtained from the following equations, which are dual to the TE equations (9.3.6):

$$\boxed{\begin{aligned}&\nabla_T^2 E_z + k_c^2 E_z = 0\\ &\boldsymbol{E}_T = -\frac{j\beta}{k_c^2}\,\boldsymbol{\nabla}_T E_z\\ &\boldsymbol{H}_T = \frac{1}{\eta_{TM}}\,\hat{\mathbf{z}}\times\boldsymbol{E}_T\end{aligned}}\qquad \text{(TM modes)}\tag{9.3.10}$$

Again, the relationship of $\boldsymbol{E}_T$ and $\boldsymbol{H}_T$ is identical to that of uniform plane waves propagating in the z-direction, but the wave impedance is now η_{TM}. The Poynting vector takes the form:

$$\mathcal{P}_z = \frac{1}{2}\,\mathrm{Re}(\boldsymbol{E}_T\times\boldsymbol{H}_T^*)\cdot\hat{\mathbf{z}} = \frac{1}{2\eta_{TM}}|\boldsymbol{E}_T|^2 = \frac{1}{2\eta_{TM}}\,\frac{\beta^2}{k_c^4}\,|\boldsymbol{\nabla}_T E_z|^2 \tag{9.3.11}$$

9.4 *Rectangular Waveguides*

Next, we discuss in detail the case of a rectangular hollow waveguide with conducting walls, as shown in Fig. 9.4.1. Without loss of generality, we may assume that the lengths a, b of the inner sides satisfy $b \le a$. The guide is typically filled with air, but any other dielectric material ϵ, μ may be assumed.

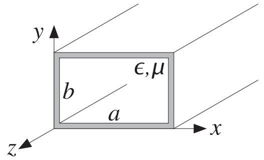


Fig. 9.4.1 Rectangular waveguide.

The simplest and dominant propagation mode is the so-called TE$_{10}$ mode and depends only on the x-coordinate (of the longest side.) Therefore, we begin by looking for solutions of Eq. (9.3.8) that depend only on x. In this case, the Helmholtz equation reduces to:

$$\partial_x^2 H_z(x) + k_c^2 H_z(x) = 0$$

The most general solution is a linear combination of $\cos k_c x$ and $\sin k_c x$. However, only the former will satisfy the boundary conditions. Therefore, the solution is:

$$H_z(x) = H_0 \cos k_c x \tag{9.4.1}$$

where H_0 is a (complex-valued) constant. Because there is no y-dependence, it follows from Eq. (9.3.8) that $\partial_y H_z = 0$, and hence $H_y = 0$ and $E_x = 0$. It also follows that:

$$H_x(x) = -\frac{j\beta}{k_c^2}\,\partial_x H_z = -\frac{j\beta}{k_c^2}(-k_c)H_0\sin k_c x = \frac{j\beta}{k_c}\,H_0\sin k_c x \equiv H_1 \sin k_c x$$

Then, the corresponding electric field will be:

$$E_y(x) = -\eta_{TE} H_x(x) = -\eta_{TE}\frac{j\beta}{k_c}\,H_0\sin k_c x \equiv E_0 \sin k_c x$$

where we defined the constants:

$$\begin{aligned} H_1 &= \frac{j\beta}{k_c}\,H_0\\ E_0 &= -\eta_{TE} H_1 = -\eta_{TE}\frac{j\beta}{k_c}\,H_0 = -j\eta\,\frac{\omega}{\omega_c}\,H_0\end{aligned} \tag{9.4.2}$$

where we used $\eta_{TE} = \eta\omega/\beta c$. In summary, the non-zero field components are:

$$\boxed{\begin{aligned}H_z(x) &= H_0\cos k_c x\\ H_x(x) &= H_1 \sin k_c x\\ E_y(x) &= E_0 \sin k_c x\end{aligned}}\quad\Rightarrow\quad \boxed{\begin{aligned}H_z(x,y,z,t) &= H_0\cos k_c x\, e^{j\omega t - j\beta z}\\ H_x(x,y,z,t) &= H_1 \sin k_c x\, e^{j\omega t - j\beta z}\\ E_y(x,y,z,t) &= E_0 \sin k_c x\, e^{j\omega t - j\beta z}\end{aligned}} \tag{9.4.3}$$

Assuming perfectly conducting walls, the boundary conditions require that there be no tangential electric field at any of the wall sides. Because the electric field is in the y-direction, it is normal to the top and bottom sides. But, it is parallel to the left and right sides. On the left side, $x = 0$, $E_y(x)$ vanishes because $\sin k_c x$ does. On the right side, $x = a$, the boundary condition requires:

$$E_y(a) = E_0 \sin k_c a = 0 \quad\Rightarrow\quad \sin k_c a = 0$$

This requires that $k_c a$ be an integral multiple of π:

$$k_c a = n\pi \quad\Rightarrow\quad k_c = \frac{n\pi}{a} \tag{9.4.4}$$

These are the so-called TE$_{n0}$ modes. The corresponding cutoff frequency $\omega_c = ck_c$, $f_c = \omega_c/2\pi$, and wavelength $\lambda_c = 2\pi/k_c = c/f_c$ are:

$$\boxed{\omega_c = \frac{cn\pi}{a}\,,\quad f_c = \frac{cn}{2a}\,,\quad \lambda_c = \frac{2a}{n}}\qquad \text{(TE}_{n0}\text{ modes)} \tag{9.4.5}$$

The dominant mode is the one with the lowest cutoff frequency or the longest cutoff wavelength, that is, the mode TE$_{10}$ having $n = 1$. It has:

$$\boxed{k_c = \frac{\pi}{a}\,,\quad \omega_c = \frac{c\pi}{a}\,,\quad f_c = \frac{c}{2a}\,,\quad \lambda_c = 2a}\qquad \text{(TE}_{10}\text{ mode)} \tag{9.4.6}$$

Fig. 9.4.2 depicts the electric field $E_y(x) = E_0 \sin k_c x = E_0 \sin(\pi x/a)$ of this mode as a function of x.

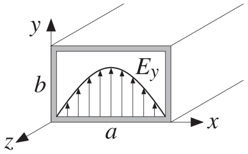


Fig. 9.4.2 Electric field inside a rectangular waveguide.

9.5 Higher TE and TM modes

To construct higher modes, we look for solutions of the Helmholtz equation that are factorable in their x and y dependence:

$$H_z(x,y)=F(x)G(y)$$

Then, Eq. (9.3.8) becomes:

$$F''(x)G(y)+F(x)G''(y)+k_c^2F(x)G(y)=0 \quad\Rightarrow\quad \frac{F''(x)}{F(x)}+\frac{G''(y)}{G(y)}+k_c^2=0 \qquad (9.5.1)$$

Because these must be valid for all x,y (inside the guide), the F- and G-terms must be constants, independent of x and y. Thus, we write:

$$\frac{F''(x)}{F(x)}=-k_x^2\,,\quad \frac{G''(y)}{G(y)}=-k_y^2 \quad \text{or}$$

$$F''(x)+k_x^2F(x)=0\,,\quad G''(y)+k_y^2G(y)=0 \qquad (9.5.2)$$

where the constants k_x^2 and k_y^2 are constrained from Eq. (9.5.1) to satisfy:

$$k_c^2=k_x^2+k_y^2 \qquad (9.5.3)$$

The most general solutions of (9.5.2) that will satisfy the TE boundary conditions are $\cos k_xx$ and $\cos k_yy$. Thus, the longitudinal magnetic field will be:

$$\boxed{H_z(x,y)=H_0\cos k_xx\cos k_yy} \qquad (\text{TE}_{nm}\text{ modes}) \qquad (9.5.4)$$

It then follows from the rest of the equations (9.3.8) that:

$$\boxed{\begin{aligned}H_x(x,y)&=H_1\sin k_xx\cos k_yy\\ H_y(x,y)&=H_2\cos k_xx\sin k_yy\end{aligned}}\qquad \boxed{\begin{aligned}E_x(x,y)&=E_1\cos k_xx\sin k_yy\\ E_y(x,y)&=E_2\sin k_xx\cos k_yy\end{aligned}} \qquad (9.5.5)$$

where we defined the constants:

$$H_1=\frac{j\beta k_x}{k_c^2}H_0\,,\quad H_2=\frac{j\beta k_y}{k_c^2}H_0$$

$$E_1=\eta_{TE}H_2=j\eta\frac{\omega k_y}{\omega_ck_c}H_0\,,\quad E_2=-\eta_{TE}H_1=-j\eta\frac{\omega k_x}{\omega_ck_c}H_0$$

The boundary conditions are that E_y vanish on the right wall, $x=a$, and that E_x vanish on the top wall, $y=b$, that is,

$$E_y(a,y)=E_{0y}\sin k_xa\cos k_yy=0\,,\quad E_x(x,b)=E_{0x}\cos k_xx\sin k_yb=0$$

The conditions require that k_xa and k_yb be integral multiples of π:

$$k_xa=n\pi\,,\quad k_yb=m\pi \quad\Rightarrow\quad \boxed{k_x=\frac{n\pi}{a}\,,\quad k_y=\frac{m\pi}{b}} \qquad (9.5.6)$$

These correspond to the TE$_{nm}$ modes. Thus, the cutoff wavenumbers of these modes $k_c=\sqrt{k_x^2+k_y^2}$ take on the quantized values:

$$\boxed{k_c=\sqrt{\left(\frac{n\pi}{a}\right)^2+\left(\frac{m\pi}{b}\right)^2}} \qquad (\text{TE}_{nm}\text{ modes}) \qquad (9.5.7)$$

The cutoff frequencies $f_{nm}=\omega_c/2\pi=ck_c/2\pi$ and wavelengths $\lambda_{nm}=c/f_{nm}$ are:

$$f_{nm}=c\sqrt{\left(\frac{n}{2a}\right)^2+\left(\frac{m}{2b}\right)^2}\,,\quad \lambda_{nm}=\frac{1}{\sqrt{\left(\frac{n}{2a}\right)^2+\left(\frac{m}{2b}\right)^2}} \qquad (9.5.8)$$

The TE$_{0m}$ modes are similar to the TE$_{n0}$ modes, but with x and a replaced by y and b. The family of TM modes can also be constructed in a similar fashion from Eq. (9.3.10).

Assuming $E_z(x,y)=F(x)G(y)$, we obtain the same equations (9.5.2). Because E_z is parallel to all walls, we must now choose the solutions $\sin k_x$ and $\sin k_yy$. Thus, the longitudinal electric fields is:

$$\boxed{E_z(x,y)=E_0\sin k_xx\sin k_yy} \qquad (\text{TM}_{nm}\text{ modes}) \qquad (9.5.9)$$

The rest of the field components can be worked out from Eq. (9.3.10) and one finds that they are given by the same expressions as (9.5.5), except now the constants are determined in terms of E_0:

$$E_1=-\frac{j\beta k_x}{k_c^2}E_0\,,\quad E_2=-\frac{j\beta k_y}{k_c^2}E_0$$

$$H_1=-\frac{1}{\eta_{TM}}E_2=\frac{j\omega k_y}{\omega_ck_c}\frac{1}{\eta}E_0\,,\quad H_2=\frac{1}{\eta_{TM}}E_1=-\frac{j\omega k_x}{\omega_ck_c}\frac{1}{\eta}H_0$$

where we used $\eta_{TM}=\eta\beta c/\omega$. The boundary conditions on E_x,E_y are the same as before, and in addition, we must require that E_z vanish on all walls.

These conditions imply that k_x,k_y will be given by Eq. (9.5.6), except both n and m must be non-zero (otherwise E_z would vanish identically.) Thus, the cutoff frequencies and wavelengths are the same as in Eq. (9.5.8).

Waveguide modes can be excited by inserting small probes at the beginning of the waveguide. The probes are chosen to generate an electric field that resembles the field of the desired mode.

9.6 Operating Bandwidth

All waveguiding systems are operated in a frequency range that ensures that only the *lowest* mode can propagate. If several modes can propagate simultaneously,[†] one has no control over which modes will actually be carrying the transmitted signal. This may cause undue amounts of dispersion, distortion, and erratic operation.

A mode with cutoff frequency ω_c will propagate only if its frequency is $\omega \geq \omega_c$, or $\lambda < \lambda_c$. If $\omega < \omega_c$, the wave will attenuate exponentially along the guide direction. This follows from the ω, β relationship (9.1.10):

$$\omega^2 = \omega_c^2 + \beta^2 c^2 \quad \Rightarrow \quad \beta^2 = \frac{\omega^2 - \omega_c^2}{c^2}$$

If $\omega \geq \omega_c$, the wavenumber β is real-valued and the wave will propagate. But if $\omega < \omega_c$, β becomes imaginary, say, $\beta = -j\alpha$, and the wave will attenuate in the z-direction, with a penetration depth $\delta = 1/\alpha$:

$$e^{-j\beta z} = e^{-\alpha z}$$

If the frequency ω is greater than the cutoff frequencies of several modes, then all of these modes can propagate. Conversely, if ω is less than all cutoff frequencies, then none of the modes can propagate.

If we arrange the cutoff frequencies in increasing order, $\omega_{c1} < \omega_{c2} < \omega_{c3} < \cdots$, then, to ensure single-mode operation, the frequency must be restricted to the interval $\omega_{c1} < \omega < \omega_{c2}$, so that only the lowest mode will propagate. This interval defines the *operating bandwidth* of the guide.

These remarks apply to all waveguiding systems, not just hollow conducting waveguides. For example, in coaxial cables the lowest mode is the TEM mode having no cutoff frequency, $\omega_{c1} = 0$. However, TE and TM modes with non-zero cutoff frequencies do exist and place an upper limit on the usable bandwidth of the TEM mode. Similarly, in optical fibers, the lowest mode has no cutoff, and the single-mode bandwidth is determined by the next cutoff frequency.

In rectangular waveguides, the smallest cutoff frequencies are $f_{10} = c/2a$, $f_{20} = c/a = 2f_{10}$, and $f_{01} = c/2b$. Because we assumed that $b \leq a$, it follows that always $f_{10} \leq f_{01}$. If $b \leq a/2$, then $1/a \leq 1/2b$ and therefore, $f_{20} \leq f_{01}$, so that the two lowest cutoff frequencies are f_{10} and f_{20}.

On the other hand, if $a/2 \leq b \leq a$, then $f_{01} \leq f_{20}$ and the two smallest frequencies are f_{10} and f_{01} (except when $b = a$, in which case $f_{01} = f_{10}$ and the smallest frequencies are f_{10} and f_{20}.) The two cases $b \leq a/2$ and $b \geq a/2$ are depicted in Fig. 9.6.1.

It is evident from this figure that in order to achieve the *widest* possible usable bandwidth for the TE$_{10}$ mode, the guide dimensions must satisfy $b \leq a/2$ so that the bandwidth is the interval $[f_c, 2f_c]$, where $f_c = f_{10} = c/2a$. In terms of the wavelength $\lambda = c/f$, the operating bandwidth becomes: $0.5 \leq a/\lambda \leq 1$, or, $a \leq \lambda \leq 2a$.

We will see later that the total amount of transmitted power in this mode is proportional to the cross-sectional area of the guide, ab. Thus, if in addition to having the widest bandwidth, we also require to have the *maximum power* transmitted, the dimension b must be chosen to be as large as possible, that is, $b = a/2$. Most practical guides follow these side proportions.

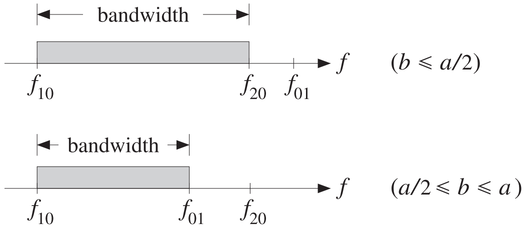


Fig. 9.6.1 Operating bandwidth in rectangular waveguides.

If there is a "canonical" guide, it will have $b = a/2$ and be operated at a frequency that lies in the middle of the operating band $[f_c, 2f_c]$, that is,

$$f = 1.5f_c = 0.75\,\frac{c}{a} \tag{9.6.1}$$

Table 9.6.1 lists some standard air-filled rectangular waveguides with their naming designations, inner side dimensions a, b in inches, cutoff frequencies in GHz, minimum and maximum recommended operating frequencies in GHz, power ratings, and attenuations in dB/m (the power ratings and attenuations are representative over each operating band.) We have chosen one example from each microwave band.

name	a	b	f_c	$f_{\min}$	$f_{\max}$	band	P	α
WR-510	5.10	2.55	1.16	1.45	2.20	L	9 MW	0.007
WR-284	2.84	1.34	2.08	2.60	3.95	S	2.7 MW	0.019
WR-159	1.59	0.795	3.71	4.64	7.05	C	0.9 MW	0.043
WR-90	0.90	0.40	6.56	8.20	12.50	X	250 kW	0.110
WR-62	0.622	0.311	9.49	11.90	18.00	Ku	140 kW	0.176
WR-42	0.42	0.17	14.05	17.60	26.70	K	50 kW	0.370
WR-28	0.28	0.14	21.08	26.40	40.00	Ka	27 kW	0.583
WR-15	0.148	0.074	39.87	49.80	75.80	V	7.5 kW	1.52
WR-10	0.10	0.05	59.01	73.80	112.00	W	3.5 kW	2.74

Table 9.6.1 Characteristics of some standard air-filled rectangular waveguides.

9.7 Power Transfer, Energy Density, and Group Velocity

Next, we calculate the time-averaged power transmitted in the TE$_{10}$ mode. We also calculate the energy density of the fields and determine the velocity by which electromagnetic energy flows down the guide and show that it is equal to the group velocity. We recall that the non-zero field components are:

$$H_z(x) = H_0 \cos k_c x\,, \quad H_x(x) = H_1 \sin k_c x\,, \quad E_y(x) = E_0 \sin k_c x \tag{9.7.1}$$

[†] Murphy's law for waveguides states that "if a mode can propagate, it will."

where

$$H_1 = \frac{j\beta}{k_c}\,H_0\,, \quad E_0 = -\eta_{TE}H_1 = -j\eta\,\frac{\omega}{\omega_c}\,H_0 \tag{9.7.2}$$

The Poynting vector is obtained from the general result of Eq. (9.3.7):

$$\mathcal{P}_z = \frac{1}{2\eta_{TE}}|\boldsymbol{E}_T|^2 = \frac{1}{2\eta_{TE}}|E_y(x)|^2 = \frac{1}{2\eta_{TE}}\,|E_0|^2\sin^2 k_cx$$

The transmitted power is obtained by integrating $\mathcal{P}_z$ over the cross-sectional area of the guide:

$$P_T = \int_0^a\int_0^b \frac{1}{2\eta_{TE}}\,|E_0|^2\sin^2 k_cx\,dxdy$$

Noting the definite integral,

$$\int_0^a \sin^2 k_cx\,dx = \int_0^a \sin^2\big(\frac{\pi x}{a}\big)dx = \frac{a}{2} \tag{9.7.3}$$

and using $\eta_{TE} = \eta\omega/\beta c = \eta/\sqrt{1-\omega_c^2/\omega^2}$, we obtain:

$$\boxed{P_T = \frac{1}{4\eta_{TE}}|E_0|^2ab = \frac{1}{4\eta}|E_0|^2ab\sqrt{1-\frac{\omega_c^2}{\omega^2}}} \quad \text{(transmitted power)} \tag{9.7.4}$$

We may also calculate the distribution of electromagnetic energy along the guide, as measured by the time-averaged energy density. The energy densities of the electric and magnetic fields are:

$$\begin{aligned} w_e &= \frac{1}{2}\,\text{Re}\big(\frac{1}{2}\epsilon\boldsymbol{E}\cdot\boldsymbol{E}^*\big) = \frac{1}{4}\epsilon|E_y|^2 \\ w_m &= \frac{1}{2}\,\text{Re}\big(\frac{1}{2}\mu\boldsymbol{H}\cdot\boldsymbol{H}^*\big) = \frac{1}{4}\mu\big(|H_x|^2+|H_z|^2\big) \end{aligned}$$

Inserting the expressions for the fields, we find:

$$w_e = \frac{1}{4}\,\epsilon|E_0|^2\sin^2 k_cx\,, \qquad w_m = \frac{1}{4}\mu\big(|H_1|^2\sin^2 k_cx + |H_0|^2\cos^2 k_cx\big)$$

Because these quantities represent the energy per unit volume, if we integrate them over the cross-sectional area of the guide, we will obtain the energy distributions per unit z-length. Using the integral (9.7.3) and an identical one for the cosine case, we find:

$$\begin{aligned} W_e' &= \int_0^a\int_0^b W_e(x,y)\,dxdy = \int_0^a\int_0^b \frac{1}{4}\,\epsilon|E_0|^2\sin^2 k_cx\,dxdy = \frac{1}{8}\,\epsilon|E_0|^2ab \\ W_m' &= \int_0^a\int_0^b \frac{1}{4}\mu\big(|H_1|^2\sin^2 k_cx + |H_0|^2\cos^2 k_cx\big)dxdy = \frac{1}{8}\mu\big(|H_1|^2+|H_0|^2\big)ab \end{aligned}$$

Although these expressions look different, they are actually equal, $W_e' = W_m'$. Indeed, using the property $\beta^2/k_c^2 + 1 = (\beta^2+k_c^2)/k_c^2 = k^2/k_c^2 = \omega^2/\omega_c^2$ and the relationships between the constants in (9.7.1), we find:

$$\mu\big(|H_1|^2+|H_0|^2\big) = \mu\big(|H_0|^2\frac{\beta^2}{k_c^2} + |H_0|^2\big) = \mu|H_0|^2\frac{\omega^2}{\omega_c^2} = \frac{\mu}{\eta^2}|E_0|^2 = \epsilon|E_0|^2$$

The equality of the electric and magnetic energies is a general property of waveguiding systems. We also encountered it in Sec. 2.3 for uniform plane waves. The total energy density per unit length will be:

$$\boxed{W' = W_e' + W_m' = 2W_e' = \frac{1}{4}\,\epsilon|E_0|^2ab} \tag{9.7.5}$$

According to the general relationship between flux, density, and transport velocity given in Eq. (1.5.2), the energy transport velocity will be the ratio $v_{\text{en}} = P_T/W'$. Using Eqs. (9.7.4) and (9.7.5) and noting that $1/\eta\epsilon = 1/\sqrt{\mu\epsilon} = c$, we find:

$$\boxed{v_{\text{en}} = \frac{P_T}{W'} = c\sqrt{1-\frac{\omega_c^2}{\omega^2}}} \quad \text{(energy transport velocity)} \tag{9.7.6}$$

This is equal to the *group velocity* of the propagating mode. For any dispersion relationship between ω and β, the group and phase velocities are defined by

$$\boxed{v_{\text{gr}} = \frac{d\omega}{d\beta}\,, \quad v_{\text{ph}} = \frac{\omega}{\beta}} \quad \text{(group and phase velocities)} \tag{9.7.7}$$

For uniform plane waves and TEM transmission lines, we have $\omega = \beta c$, so that $v_{\text{gr}} = v_{\text{ph}} = c$. For a rectangular waveguide, we have $\omega^2 = \omega_c^2 + \beta^2c^2$. Taking differentials of both sides, we find $2\omega d\omega = 2c^2\beta d\beta$, which gives:

$$\boxed{v_{\text{gr}} = \frac{d\omega}{d\beta} = \frac{\beta c^2}{\omega} = c\sqrt{1-\frac{\omega_c^2}{\omega^2}}} \tag{9.7.8}$$

where we used Eq. (9.1.10). Thus, the energy transport velocity is equal to the group velocity, $v_{\text{en}} = v_{\text{gr}}$. We note that $v_{\text{gr}} = \beta c^2/\omega = c^2/v_{\text{ph}}$, or

$$v_{\text{gr}}v_{\text{ph}} = c^2 \tag{9.7.9}$$

The energy or group velocity satisfies $v_{\text{gr}} \le c$, whereas $v_{\text{ph}} \ge c$. Information transmission down the guide is by the group velocity and, consistent with the theory of relativity, it is less than c.

9.8 Power Attenuation

In this section, we calculate the attenuation coefficient due to the ohmic losses of the conducting walls following the procedure outlined in Sec. 9.2. The losses due to the filling dielectric can be determined from Eq. (9.2.5).

The field expressions (9.4.3) were derived assuming the boundary conditions for perfectly conducting wall surfaces. The induced surface currents on the inner walls of the waveguide are given by $\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H}$, where the unit vector $\hat{\mathbf{n}}$ is $\pm\hat{\mathbf{x}}$ and $\pm\hat{\mathbf{y}}$ on the left/right and bottom/top walls, respectively.

The surface currents and tangential magnetic fields are shown in Fig. 9.8.1. In particular, on the bottom and top walls, we have:

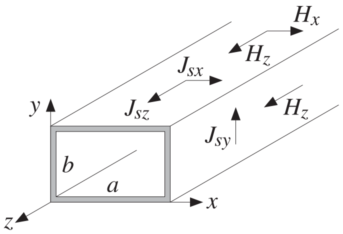


Fig. 9.8.1 Currents on waveguide walls.

$$\boldsymbol{J}_s = \pm\hat{\mathbf{y}}\times\boldsymbol{H} = \pm\hat{\mathbf{y}}\times(\hat{\mathbf{x}}H_x+\hat{\mathbf{z}}H_z) = \pm(-\hat{\mathbf{z}}H_x+\hat{\mathbf{x}}H_z) = \pm(-\hat{\mathbf{z}}H_1\sin k_cx+\hat{\mathbf{x}}H_0\cos k_cx)$$

Similarly, on the left and right walls:

$$\boldsymbol{J}_s = \pm\hat{\mathbf{x}}\times\boldsymbol{H} = \pm\hat{\mathbf{x}}\times(\hat{\mathbf{x}}H_x+\hat{\mathbf{z}}H_z) = \mp\hat{\mathbf{y}}H_z = \mp\hat{\mathbf{y}}H_0\cos k_cx$$

At $x = 0$ and $x = a$, this gives $\boldsymbol{J}_s = \mp\hat{\mathbf{y}}(\pm H_0) = \hat{\mathbf{y}}H_0$. Thus, the magnitudes of the surface currents are on the four walls:

$$|\boldsymbol{J}_s|^2 = \begin{cases} |H_0|^2\,, & \text{(left and right walls)}\\ |H_0|^2\cos^2 k_cx + |H_1|^2\sin^2 k_cx\,, & \text{(top and bottom walls)}\end{cases}$$

The power loss per unit z-length is obtained from Eq. (9.2.8) by integrating $|\boldsymbol{J}_s|^2$ around the four walls, that is,

$$\begin{aligned} P'_{\text{loss}} &= 2\frac{1}{2}R_s\int_0^a |\boldsymbol{J}_s|^2\,dx + 2\frac{1}{2}R_s\int_0^b|\boldsymbol{J}_s|^2\,dy\\ &= R_s\int_0^a\big(|H_0|^2\cos^2 k_cx + |H_1|^2\sin^2 k_cx\big)\,dx + R_s\int_0^b |H_0|^2\,dy\\ &= R_s\frac{a}{2}\big(|H_0|^2+|H_1|^2\big) + R_sb|H_0|^2 = \frac{R_sa}{2}\Big(|H_0|^2+|H_1|^2+\frac{2b}{a}|H_0|^2\Big)\end{aligned}$$

Using $|H_0|^2+|H_1|^2 = |E_0|^2/\eta^2$ from Sec. 9.7, and $|H_0|^2 = (|E_0|^2/\eta^2)\,\omega_c^2/\omega^2$, which follows from Eq. (9.4.2), we obtain:

$$P'_{\text{loss}} = \frac{R_sa|E_0|^2}{2\eta^2}\Big(1+\frac{2b}{a}\frac{\omega_c^2}{\omega^2}\Big)$$

The attenuation constant is computed from Eqs. (9.2.9) and (9.7.4):

$$\alpha_c = \frac{P'_{\text{loss}}}{2P_T} = \frac{\dfrac{R_sa|E_0|^2}{2\eta^2}\Big(1+\dfrac{2b}{a}\dfrac{\omega_c^2}{\omega^2}\Big)}{2\dfrac{1}{4\eta}|E_0|^2ab\sqrt{1-\dfrac{\omega_c^2}{\omega^2}}}$$

which gives:

$$\boxed{\alpha_c = \frac{R_s}{\eta b}\,\frac{\Big(1+\dfrac{2b}{a}\dfrac{\omega_c^2}{\omega^2}\Big)}{\sqrt{1-\dfrac{\omega_c^2}{\omega^2}}}}\qquad\text{(attenuation of TE}_{10}\text{ mode)}\qquad(9.8.1)$$

This is in units of nepers/m. Its value in dB/m is obtained by $\alpha_{\text{dB}} = 8.686\alpha_c$. For a given ratio a/b, α_c increases with decreasing b, thus the smaller the guide dimensions, the larger the attenuation. This trend is noted in Table 9.6.1.

The main tradeoffs in a waveguiding system are that as the operating frequency f increases, the dimensions of the guide must decrease in order to maintain the operating band $f_c \le f \le 2f_c$, but then the attenuation increases and the transmitted power decreases as it is proportional to the guide's area.

Example 9.8.1: Design a rectangular air-filled waveguide to be operated at 5 GHz, then, redesign it to be operated at 10 GHz. The operating frequency must lie in the middle of the operating band. Calculate the guide dimensions, the attenuation constant in dB/m, and the maximum transmitted power assuming the maximum electric field is one-half of the dielectric strength of air. Assume copper walls with conductivity $\sigma = 5.8\times10^7$ S/m.

Solution: If f is in the middle of the operating band, $f_c \le f \le 2f_c$, where $f_c = c/2a$, then $f = 1.5f_c = 0.75c/a$. Solving for a, we find

$$a = \frac{0.75c}{f} = \frac{0.75\times30\text{ GHz cm}}{5} = 4.5\text{ cm}$$

For maximum power transfer, we require $b = a/2 = 2.25$ cm. Because $\omega = 1.5\omega_c$, we have $\omega_c/\omega = 2/3$. Then, Eq. (9.8.1) gives $\alpha_c = 0.037$ dB/m. The dielectric strength of air is 3 MV/m. Thus, the maximum allowed electric field in the guide is $E_0 = 1.5$ MV/m. Then, Eq. (9.7.4) gives $P_T = 1.12$ MW.

At 10 GHz, because f is doubled, the guide dimensions are halved, $a = 2.25$ and $b = 1.125$ cm. Because R_s depends on f like $f^{1/2}$, it will increase by a factor of $\sqrt{2}$. Then, the factor R_s/b will increase by a factor of $2\sqrt{2}$. Thus, the attenuation will increase to the value $\alpha_c = 0.037\cdot2\sqrt{2} = 0.104$ dB/m. Because the area ab is reduced by a factor of four, so will the power, $P_T = 1.12/4 = 0.28$ MW $=$ 280 kW.

The results of these two cases are consistent with the values quoted in Table 9.6.1 for the C-band and X-band waveguides, WR-159 and WR-90. □

Example 9.8.2: *WR-159 Waveguide.* Consider the C-band WR-159 air-filled waveguide whose characteristics were listed in Table 9.6.1. Its inner dimensions are $a = 1.59$ and $b = a/2 = 0.795$ inches, or, equivalently, $a = 4.0386$ and $b = 2.0193$ cm.

The cutoff frequency of the TE_{10} mode is $f_c = c/2a = 3.71$ GHz. The maximum operating bandwidth is the interval $[f_c, 2f_c] = [3.71, 7.42]$ GHz, and the recommended interval is $[4.64, 7.05]$ GHz.

Assuming copper walls with conductivity $\sigma = 5.8\times10^7$ S/m, the calculated attenuation constant α_c from Eq. (9.8.1) is plotted in dB/m versus frequency in Fig. 9.8.2.

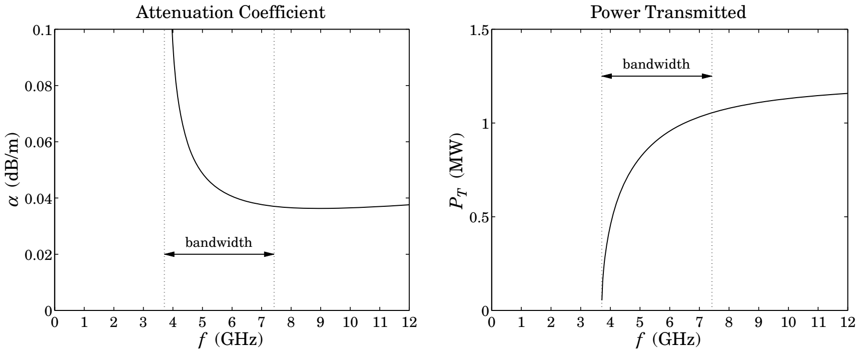


Fig. 9.8.2 Attenuation constant and transmitted power in a WR-159 waveguide.

The power transmitted P_T is calculated from Eq. (9.7.4) assuming a maximum breakdown voltage of $E_0 = 1.5$ MV/m, which gives a safety factor of two over the dielectric breakdown of air of 3 MV/m. The power in megawatt scales is plotted in Fig. 9.8.2.

Because of the factor $\sqrt{1-\omega_c^2/\omega^2}$ in the denominator of α_c and the numerator of P_T, the attenuation constant becomes very large near the cutoff frequency, while the power is almost zero. A physical explanation of this behavior is given in the next section. □

9.9 Reflection Model of Waveguide Propagation

An intuitive model for the TE_{10} mode can be derived by considering a TE-polarized uniform plane wave propagating in the z-direction by obliquely bouncing back and forth between the left and right walls of the waveguide, as shown in Fig. 9.9.1.

If θ is the angle of incidence, then the incident and reflected (from the right wall) wavevectors will be:

$$\begin{aligned}\mathbf{k} &= \hat{\mathbf{x}}\,k_x + \hat{\mathbf{z}}\,k_z = \hat{\mathbf{x}}\,k\cos\theta + \hat{\mathbf{z}}\,k\sin\theta \\ \mathbf{k}' &= -\hat{\mathbf{x}}\,k_x + \hat{\mathbf{z}}\,k_z = -\hat{\mathbf{x}}\,k\cos\theta + \hat{\mathbf{z}}\,k\sin\theta\end{aligned}$$

The electric and magnetic fields will be the sum of an incident and a reflected component of the form:

$$\begin{aligned}\mathbf{E} &= \hat{\mathbf{y}}\,E_1 e^{-j\mathbf{k}\cdot\mathbf{r}} + \hat{\mathbf{y}}\,E_1' e^{-j\mathbf{k}'\cdot\mathbf{r}} = \hat{\mathbf{y}}\,E_1 e^{-jk_x x}e^{-jk_z z} + \hat{\mathbf{y}}\,E_1' e^{jk_x x}e^{-jk_z z} = \mathbf{E}_1 + \mathbf{E}_1' \\ \mathbf{H} &= \frac{1}{\eta}\hat{\mathbf{k}}\times\mathbf{E}_1 + \frac{1}{\eta}\hat{\mathbf{k}}'\times\mathbf{E}_1'\end{aligned}$$

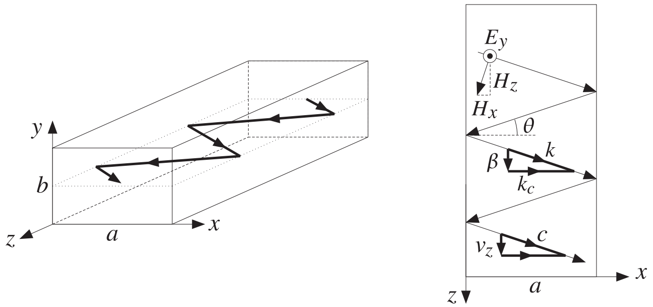


Fig. 9.9.1 Reflection model of TE_{10} mode.

where the electric field was taken to be polarized in the y direction. These field expressions become component-wise:

$$\begin{aligned}E_y &= \left(E_1 e^{-jk_x x} + E_1' e^{jk_x x}\right)e^{-jk_z z} \\ H_x &= -\frac{1}{\eta}\sin\theta\left(E_1 e^{-jk_x x} + E_1' e^{jk_x x}\right)e^{-jk_z z} \\ H_z &= \frac{1}{\eta}\cos\theta\left(E_1 e^{-jk_x x} - E_1' e^{jk_x x}\right)e^{-jk_z z}\end{aligned} \tag{9.9.1}$$

The boundary condition on the left wall, $x = 0$, requires that $E_1 + E_1' = 0$. We may write therefore, $E_1 = -E_1' = jE_0/2$. Then, the above expressions simplify into:

$$\begin{aligned}E_y &= E_0\sin k_x x\, e^{-jk_z z} \\ H_x &= -\frac{1}{\eta}\sin\theta E_0\sin k_x x\, e^{-jk_z z} \\ H_z &= \frac{j}{\eta}\cos\theta E_0\cos k_x x\, e^{-jk_z z}\end{aligned} \tag{9.9.2}$$

These are identical to Eq. (9.4.3) provided we identify β with k_z and k_c with k_x, as shown in Fig. 9.9.1. It follows from the wavevector triangle in the figure that the angle of incidence θ will be given by $\cos\theta = k_x/k = k_c/k$, or,

$$\boxed{\cos\theta = \frac{\omega_c}{\omega}\,,\quad \sin\theta = \sqrt{1-\frac{\omega_c^2}{\omega^2}}} \tag{9.9.3}$$

The ratio of the transverse components, $-E_y/H_x$, is the transverse impedance, which is recognized to be η_{TE}. Indeed, we have:

$$\eta_{TE} = -\frac{E_y}{H_x} = \frac{\eta}{\sin\theta} = \frac{\eta}{\sqrt{1-\dfrac{\omega_c^2}{\omega^2}}} \tag{9.9.4}$$

The boundary condition on the right wall requires $\sin k_x a = 0$, which gives rise to the same condition as (9.4.4), that is, $k_c a = n\pi$.

This model clarifies also the meaning of the group velocity. The plane wave is bouncing left and right with the speed of light c. However, the component of this velocity in the z-direction will be $v_z = c\sin\theta$. This is equal to the group velocity. Indeed, it follows from Eq. (9.9.3) that:

$$v_z = c\sin\theta = c\sqrt{1-\frac{\omega_c^2}{\omega^2}} = v_{\text{gr}} \tag{9.9.5}$$

Eq. (9.9.3) implies also that at $\omega = \omega_c$, we have $\sin\theta = 0$, or $\theta = 0$, that is, the wave is bouncing left and right at normal incidence, creating a standing wave, and does not propagate towards the z-direction. Thus, the transmitted power is zero and this also implies, through Eq. (9.2.9), that α_c will be infinite.

On the other hand, for very large frequencies, $\omega \gg \omega_c$, the angle θ will tend to 90°, causing the wave to zoom through guide almost at the speed of light.

9.10 Resonant Cavities

Cavity resonators are metallic enclosures that can trap electromagnetic fields. The boundary conditions on the cavity walls force the fields to exist only at certain quantized resonant frequencies. For highly conducting walls, the resonances are extremely sharp, having a very high Q of the order of 10,000.

Because of their high Q, cavities can be used not only to efficiently store electromagnetic energy at microwave frequencies, but also to act as precise oscillators and to perform precise frequency measurements.

Fig. 9.10.1 shows a rectangular cavity with z-length equal to l formed by replacing the sending and receiving ends of a waveguide by metallic walls. A forward-moving wave will bounce back and forth from these walls, resulting in a standing-wave pattern along the z-direction.

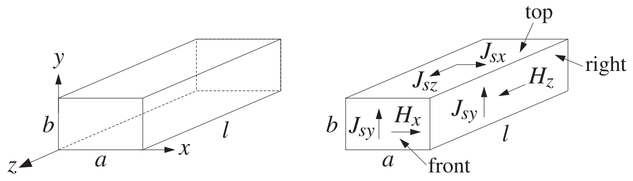


Fig. 9.10.1 Rectangular cavity resonator (and induced wall currents for the TE_{n0p} mode.)

Because the tangential components of the electric field must vanish at the end-walls, these walls must coincide with zero crossings of the standing wave, or put differently, an integral multiple of half-wavelengths must fit along the z-direction, that is, $l = p\lambda_g/2 = p\pi/\beta$, or $\beta = p\pi/l$, where p is a non-zero integer. For the same reason, the standing-wave patterns along the transverse directions require $a = n\lambda_x/2$ and $b = m\lambda_y/2$, or $k_x = n\pi/a$ and $k_y = m\pi/b$. Thus, all three cartesian components of the wave vector are quantized, and therefore, so is the frequency of the wave $\omega = c\sqrt{k_x^2+k_y^2+\beta^2}$:

$$\boxed{\omega_{nmp} = c\sqrt{\left(\frac{n\pi}{a}\right)^2+\left(\frac{m\pi}{b}\right)^2+\left(\frac{p\pi}{l}\right)^2}} \quad \text{(resonant frequencies)} \tag{9.10.1}$$

Such modes are designated as TE_{nmp} or TM_{nmp}. For simplicity, we consider the case TE_{n0p}. Eqs. (9.3.6) also describe backward-moving waves if one replaces β by $-\beta$, which also changes the sign of $\eta_{\text{TE}} = \eta\omega/\beta c$. Starting with a linear combination of forward and backward waves in the TE_{n0} mode, we obtain the field components:

$$\begin{aligned}
H_z(x,z) &= H_0\cos k_c x\,(Ae^{-j\beta z}+Be^{j\beta z}),\\
H_x(x,z) &= jH_1\sin k_c x\,(Ae^{-j\beta z}-Be^{j\beta z}), \quad H_1 = \frac{\beta}{k_c}H_0\\
E_y(x,z) &= -jE_0\sin k_c x\,(Ae^{-j\beta z}+Be^{j\beta z}), \quad E_0 = \frac{\omega}{\omega_c}\eta H_0
\end{aligned} \tag{9.10.2}$$

where $\omega_c = ck_c$. By requiring that $E_y(x,z)$ have z-dependence of the form $\sin\beta z$, the coefficients A, B must be chosen as $A = -B = j/2$. Then, Eq. (9.10.2) specializes into:

$$\begin{aligned}
H_z(x,z) &= H_0\cos k_c x\,\sin\beta z\,,\\
H_x(x,z) &= -H_1\sin k_c x\,\cos\beta z\,, \quad H_1 = \frac{\beta}{k_c}H_0\\
E_y(x,z) &= -jE_0\sin k_c x\,\sin\beta z\,, \quad E_0 = \frac{\omega}{\omega_c}\eta H_0
\end{aligned} \tag{9.10.3}$$

As expected, the vanishing of $E_y(x,z)$ on the front/back walls, $z = 0$ and $z = l$, and on the left/right walls, $x = 0$ and $x = a$, requires the quantization conditions: $\beta = p\pi/l$ and $k_c = n\pi/a$. The Q of the resonator can be calculated from its definition:

$$Q = \omega\frac{W}{P_{\text{loss}}} \tag{9.10.4}$$

where W is the total time-averaged energy stored within the cavity volume and P_{loss} is the total power loss due to the wall ohmic losses (plus other losses, such as dielectric losses, if present.) The ratio $\Delta\omega = P_{\text{loss}}/W$ is usually identified as the 3-dB width of the resonance centered at frequency ω. Therefore, we may write $Q = \omega/\Delta\omega$.

It is easily verified that the electric and magnetic energies are equal, therefore, W may be calculated by integrating the electric energy density over the cavity volume:

$$\begin{aligned}
W = 2W_e &= 2\frac{1}{4}\int_{\text{vol}}\epsilon|E_y(x,z)|^2\,dx\,dy\,dz = \frac{1}{2}\epsilon|E_0|^2\int_0^a\int_0^b\int_0^l \sin^2 k_c x\,\cos^2\beta z\,dx\,dy\,dz\\
&= \frac{1}{8}\epsilon|E_0|^2(abl) = \frac{1}{8}\mu|H_0|^2\frac{\omega^2}{\omega_c^2}(abl) = \frac{1}{8}\mu\,|H_0|^2\left[\frac{k_c^2+\beta^2}{k_c^2}\right](abl)
\end{aligned}$$

where we used the following definite integrals (valid because $k_c = n\pi/a$, $\beta = p\pi/l$):

$$\int_0^a\sin^2 k_c x\,dx = \int_0^a\cos^2 k_c x\,dx = \frac{a}{2}\,, \quad \int_0^l\sin^2\beta z\,dz = \int_0^l\cos^2\beta z\,dz = \frac{l}{2} \tag{9.10.5}$$

The ohmic losses are calculated from Eq. (9.2.6), integrated over all six cavity sides. The surface currents induced on the walls are related to the tangential magnetic fields by $\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H}_{\text{tan}}$. The directions of these currents are shown in Fig. 9.10.1. Specifically, we find for the currents on the six sides:

$$|\boldsymbol{J}_s|^2 = \begin{cases} H_0^2\sin^2\beta z & \text{(left \& right)}\\ H_0^2\cos^2 k_c x\,\sin^2\beta z + H_1^2\sin^2 k_c x\,\cos^2\beta z & \text{(top \& bottom)}\\ H_1^2\sin^2 k_c x & \text{(front \& back)}\end{cases}$$

The power loss can be computed by integrating the loss per unit conductor area, Eq. (9.2.6), over the six wall sides, or doubling the answer for the left, top, and front sides. Using the integrals (9.10.5), we find:

$$\begin{aligned} P_{\text{loss}} &= \frac{1}{2}R_s\int_{\text{walls}}|\boldsymbol{J}_s|^2\,dA = R_s\left[H_0^2\frac{bl}{2} + (H_0^2+H_1^2)\frac{al}{4} + H_1^2\frac{ab}{2}\right]\\ &= \frac{1}{4}R_s H_0^2\left[l(2b+a) + \frac{\beta^2}{k_c^2}a(2b+l)\right]\end{aligned}\tag{9.10.6}$$

where we substituted $H_1^2 = H_0^2\beta^2/k_c^2$. It follows that the Q-factor will be:

$$Q = \omega\,\frac{W}{P_{\text{loss}}} = \frac{\omega\mu}{2R_s}\,\frac{(k_c^2+\beta^2)\,(abl)}{k_c^2 l(2b+a) + \beta^2 a(2b+l)}$$

For the TE$_{n0p}$ mode we have $\beta = p\pi/l$ and $k_c = n\pi/a$. Using Eq. (9.2.7) to replace R_s in terms of the skin depth δ, we find:

$$Q = \frac{1}{\delta}\,\frac{\dfrac{n^2}{a^2} + \dfrac{p^2}{l^2}}{\dfrac{n^2}{a^2}\left(\dfrac{2}{a}+\dfrac{1}{b}\right) + \dfrac{p^2}{l^2}\left(\dfrac{2}{l}+\dfrac{1}{b}\right)}\tag{9.10.7}$$

The lowest resonant frequency corresponds to $n = p = 1$. For a cubic cavity, $a = b = l$, the Q and the lowest resonant frequency are:

$$Q = \frac{a}{3\delta}\,,\quad \omega_{101} = \frac{c\pi\sqrt{2}}{a}\,,\quad f_{101} = \frac{\omega}{2\pi} = \frac{c}{a\sqrt{2}}\tag{9.10.8}$$

For an air-filled cubic cavity with $a = 3$ cm, we find $f_{101} = 7.07$ GHz, $\delta = 7.86\times 10^{-5}$ cm, and $Q = 12724$. As in waveguides, cavities can be excited by inserting small probes that generate fields resembling a particular mode.

9.11 Dielectric Slab Waveguides

A dielectric slab waveguide is a planar dielectric sheet or thin film of some thickness, say $2a$, as shown in Fig. 9.11.1. Wave propagation in the z-direction is by total internal reflection from the left and right walls of the slab. Such waveguides provide simple models for the confining mechanism of waves propagating in optical fibers.

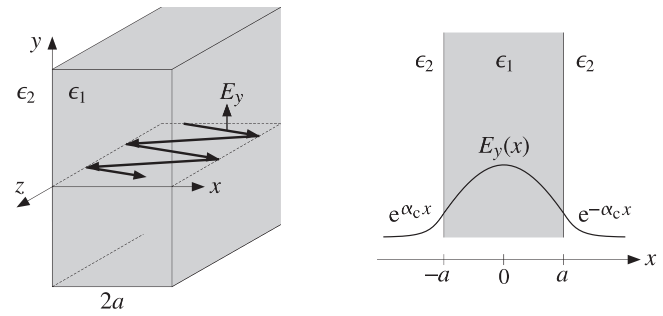


Fig. 9.11.1 Dielectric slab waveguide.

The propagating fields are confined primarily inside the slab, however, they also exist as evanescent waves outside it, decaying exponentially with distance from the slab. Fig. 9.11.1 shows a typical electric field pattern as a function of x.

For simplicity, we assume that the media to the left and right of the slab are the same. To guarantee total internal reflection, the dielectric constants inside and outside the slab must satisfy $\epsilon_1 > \epsilon_2$, and similarly for the refractive indices, $n_1 > n_2$.

We look for TE solutions that depend only on the x coordinate. The cutoff wavenumber k_c appearing in the Helmholtz equation for $H_z(x)$ depends on the dielectric constant of the propagation medium, $k_c^2 = \omega^2\epsilon\mu - \beta^2$. Therefore, k_c^2 takes different values inside and outside the guide:

$$\begin{aligned} k_{c1}^2 &= \omega^2\epsilon_1\mu_0 - \beta^2 = \omega^2\epsilon_0\mu_0 n_1^2 - \beta^2 = k_0^2 n_1^2 - \beta^2 \qquad \text{(inside)}\\ k_{c2}^2 &= \omega^2\epsilon_2\mu_0 - \beta^2 = \omega^2\epsilon_0\mu_0 n_2^2 - \beta^2 = k_0^2 n_2^2 - \beta^2 \qquad \text{(outside)}\end{aligned}\tag{9.11.1}$$

where $k_0 = \omega/c_0$ is the free-space wavenumber. We note that ω, β are the same inside and outside the guide. This follows from matching the tangential fields at all times t and all points z along the slab walls. The corresponding Helmholtz equations in the regions inside and outside the guide are:

$$\begin{aligned} \partial_x^2 H_z(x) + k_{c1}^2 H_z(x) &= 0 \quad \text{for} \quad |x| \le a\\ \partial_x^2 H_z(x) + k_{c2}^2 H_z(x) &= 0 \quad \text{for} \quad |x| \ge a\end{aligned}\tag{9.11.2}$$

Inside the slab, the solutions are $\sin k_{c1}x$ and $\cos k_{c1}x$, and outside, $\sin k_{c2}x$ and $\cos k_{c2}x$, or equivalently, $e^{\pm jk_{c2}x}$. In order for the waves to remain confined in the near vicinity of the slab, the quantity k_{c2} must be *imaginary*, for if it is real, the fields would propagate at large x distances from the slab (they would correspond to the rays refracted from the inside into the outside.)

If we set $k_{c2} = -j\alpha_c$, the solutions outside will be $e^{\pm\alpha_c x}$. If α_c is positive, then only the solution $e^{-\alpha_c x}$ is physically acceptable to the right of the slab, $x \ge a$, and only $e^{\alpha_c x}$ to the left, $x \le -a$. Thus, the fields attenuate exponentially with the transverse distance x, and exist effectively within a skin depth distance $1/\alpha_c$ from the slab. Setting $k_{c1} = k_c$ and $k_{c2} = -j\alpha_c$, Eqs. (9.11.1) become in this new notation:

$$\boxed{\begin{aligned} k_c^2 &= k_0^2 n_1^2 - \beta^2 \\ -\alpha_c^2 &= k_0^2 n_2^2 - \beta^2 \end{aligned}} \quad \Rightarrow \quad \boxed{\begin{aligned} k_c^2 &= k_0^2 n_1^2 - \beta^2 \\ \alpha_c^2 &= \beta^2 - k_0^2 n_2^2 \end{aligned}} \tag{9.11.3}$$

Similarly, Eqs. (9.11.2) read:

$$\begin{aligned} \partial_x^2 H_z(x) + k_c^2 H_z(x) &= 0 \quad \text{for} \quad |x| \le a \\ \partial_x^2 H_z(x) - \alpha_c^2 H_z(x) &= 0 \quad \text{for} \quad |x| \ge a \end{aligned} \tag{9.11.4}$$

The two solutions $\sin k_c x$ and $\cos k_c x$ inside the guide give rise to the so-called even and odd TE modes (referring to the even-ness or oddness of the resulting electric field.) For the *even modes*, the solutions of Eqs. (9.11.4) have the form:

$$H_z(x) = \begin{cases} H_1 \sin k_c x\,, & \text{if} \quad -a \le x \le a \\ H_2 e^{-\alpha_c x}\,, & \text{if} \quad x \ge a \\ H_3 e^{\alpha_c x}\,, & \text{if} \quad x \le -a \end{cases} \tag{9.11.5}$$

The corresponding x-components H_x are obtained by applying Eq. (9.3.8) using the appropriate value for k_c^2, that is, $k_{c2}^2 = -\alpha_c^2$ outside and $k_{c1}^2 = k_c^2$ inside:

$$H_x(x) = \begin{cases} -\dfrac{j\beta}{k_c^2}\partial_x H_z(x) = -\dfrac{j\beta}{k_c} H_1 \cos k_c x\,, & \text{if} \quad -a \le x \le a \\ -\dfrac{j\beta}{-\alpha_c^2}\partial_x H_z(x) = -\dfrac{j\beta}{\alpha_c} H_2 e^{-\alpha_c x}\,, & \text{if} \quad x \ge a \\ -\dfrac{j\beta}{-\alpha_c^2}\partial_x H_z(x) = \dfrac{j\beta}{\alpha_c} H_3 e^{\alpha_c x}\,, & \text{if} \quad x \ge a \end{cases} \tag{9.11.6}$$

The electric fields are $E_y(x) = -\eta_{TE} H_x(x)$, where $\eta_{TE} = \omega\mu_0/\beta$ is the *same* inside and outside the slab. Thus, the electric field has the form:

$$E_y(x) = \begin{cases} E_1 \cos k_c x\,, & \text{if} \quad -a \le x \le a \\ E_2 e^{-\alpha_c x}\,, & \text{if} \quad x \ge a \\ E_3 e^{\alpha_c x}\,, & \text{if} \quad x \le -a \end{cases} \qquad \text{(even TE modes)} \tag{9.11.7}$$

where we defined the constants:

$$E_1 = \frac{j\beta}{k_c}\eta_{TE} H_1\,, \quad E_2 = \frac{j\beta}{\alpha_c}\eta_{TE} H_2\,, \quad E_3 = -\frac{j\beta}{\alpha_c}\eta_{TE} H_3 \tag{9.11.8}$$

The boundary conditions state that the tangential components of the magnetic and electric fields, that is, H_z, E_y, are continuous across the dielectric interfaces at $x = -a$ and $x = a$. Similarly, the normal components of the magnetic field $B_x = \mu_0 H_x$ and therefore also H_x must be continuous. Because $E_y = -\eta_{TE} H_x$ and η_{TE} is the same in both media, the continuity of E_y follows from the continuity of H_x. The continuity of H_z at $x = a$ and $x = -a$ implies that:

$$H_1 \sin k_c a = H_2 e^{-\alpha_c a} \quad \text{and} \quad -H_1 \sin k_c a = H_3 e^{-\alpha_c a} \tag{9.11.9}$$

Similarly, the continuity of H_x implies (after canceling a factor of $-j\beta$):

$$\frac{1}{k_c} H_1 \cos k_c a = \frac{1}{\alpha_c} H_2 e^{-\alpha_c a} \quad \text{and} \quad \frac{1}{k_c} H_1 \cos k_c a = -\frac{1}{\alpha_c} H_3 e^{-\alpha_c a} \tag{9.11.10}$$

Eqs. (9.11.9) and (9.11.10) imply:

$$\boxed{H_2 = -H_3 = H_1 e^{\alpha_c a} \sin k_c a = H_1 e^{\alpha_c a} \frac{\alpha_c}{k_c} \cos k_c a} \tag{9.11.11}$$

Similarly, we find for the electric field constants:

$$\boxed{E_2 = E_3 = E_1 e^{\alpha_c a} \cos k_c a = E_1 e^{\alpha_c a} \frac{k_c}{\alpha_c} \sin k_c a} \tag{9.11.12}$$

The consistency of the last equations in (9.11.11) or (9.11.12) requires that:

$$\cos k_c a = \frac{k_c}{\alpha_c} \sin k_c a \quad \Rightarrow \quad \boxed{\alpha_c = k_c \tan k_c a} \tag{9.11.13}$$

For the *odd* TE modes, we have for the solutions of Eq. (9.11.4):

$$H_z(x) = \begin{cases} H_1 \cos k_c x\,, & \text{if} \quad -a \le x \le a \\ H_2 e^{-\alpha_c x}\,, & \text{if} \quad x \ge a \\ H_3 e^{\alpha_c x}\,, & \text{if} \quad x \le -a \end{cases} \tag{9.11.14}$$

The resulting electric field is:

$$E_y(x) = \begin{cases} E_1 \sin k_c x\,, & \text{if} \quad -a \le x \le a \\ E_2 e^{-\alpha_c x}\,, & \text{if} \quad x \ge a \\ E_3 e^{\alpha_c x}\,, & \text{if} \quad x \le -a \end{cases} \qquad \text{(odd TE modes)} \tag{9.11.15}$$

The boundary conditions imply in this case:

$$\boxed{H_2 = H_3 = H_1 e^{\alpha_c a} \cos k_c a = -H_1 e^{\alpha_c a} \frac{\alpha_c}{k_c} \sin k_c a} \tag{9.11.16}$$

and, for the electric field constants:

$$\boxed{E_2 = -E_3 = E_1 e^{\alpha_c a} \sin k_c a = -E_1 e^{\alpha_c a} \frac{k_c}{\alpha_c} \cos k_c a} \tag{9.11.17}$$

The consistency of the last equation requires:

$$\boxed{\alpha_c = -k_c \cot k_c a} \tag{9.11.18}$$

We note that the electric fields $E_y(x)$ given by Eqs. (9.11.7) and (9.11.15) are even or odd functions of x for the two families of modes. Expressing E_2 and E_3 in terms of E_1, we summarize the forms of the electric fields in the two cases:

$$E_y(x)=\begin{cases} E_1\cos k_c x\,, & \text{if}\quad -a\le x\le a\\ E_1\cos k_c a\, e^{-\alpha_c(x-a)}\,, & \text{if}\quad x\ge a\\ E_1\cos k_c a\, e^{\alpha_c(x+a)}\,, & \text{if}\quad x\le -a\end{cases}\qquad\text{(even TE modes)}\qquad(9.11.19)$$

$$E_y(x)=\begin{cases} E_1\sin k_c x\,, & \text{if}\quad -a\le x\le a\\ E_1\sin k_c a\, e^{-\alpha_c(x-a)}\,, & \text{if}\quad x\ge a\\ -E_1\sin k_c a\, e^{\alpha_c(x+a)}\,, & \text{if}\quad x\le -a\end{cases}\qquad\text{(odd TE modes)}\qquad(9.11.20)$$

Given the operating frequency ω, Eqs. (9.11.3) and (9.11.13) or (9.11.18) provide three equations in the three unknowns k_c, α_c, β. To solve them, we add the two equations (9.11.3) to eliminate β:

$$\boxed{\alpha_c^2+k_c^2=k_0^2(n_1^2-n_2^2)=\frac{\omega^2}{c_0^2}(n_1^2-n_2^2)}\qquad(9.11.21)$$

Next, we discuss the numerical solutions of these equations. Defining the dimensionless quantities $u=k_c a$ and $v=\alpha_c a$, we may rewrite Eqs. (9.11.13), (9.11.18), and (9.11.21) in the equivalent forms:

$$\boxed{\begin{matrix} v=u\tan u\\ v^2+u^2=R^2\end{matrix}\quad\text{(even modes)}}\,,\qquad \boxed{\begin{matrix} v=-u\cot u\\ v^2+u^2=R^2\end{matrix}\quad\text{(odd modes)}}\qquad(9.11.22)$$

where R is the normalized frequency variable:

$$\boxed{R=k_0 a N_A=\frac{\omega a}{c_0}N_A=\frac{2\pi f a}{c_0}N_A=\frac{2\pi a}{\lambda}N_A}\qquad(9.11.23)$$

where $N_A=\sqrt{n_1^2-n_2^2}$ is the *numerical aperture* of the slab and $\lambda=c_0/f$, the free-space wavelength.

Because the functions $\tan u$ and $\cot u$ have many branches, there may be several possible solution pairs u, v for each value of R. These solutions are obtained at the intersections of the curves $v=u\tan u$ and $v=-u\cot u$ with the circle of radius R, that is, $v^2+u^2=R^2$. Fig. 9.11.2 shows the solutions for various values of the radius R corresponding to various values of ω.

It is evident from the figure that for small enough R, that is, $0\le R<\pi/2$, there is only one solution and it is even (for an optical fiber, the single-mode condition reads $2\pi a N_A/\lambda<2.405$, where a is the core radius.) For $\pi/2\le R<\pi$, there are two solutions, one even and one odd. For $\pi\le R<3\pi/2$, there are three solutions, two even and one odd, and so on. In general, there will be $M+1$ solutions, alternating between even and odd, if R falls in the interval:

$$\frac{M\pi}{2}\le R<\frac{(M+1)\pi}{2}\qquad(9.11.24)$$

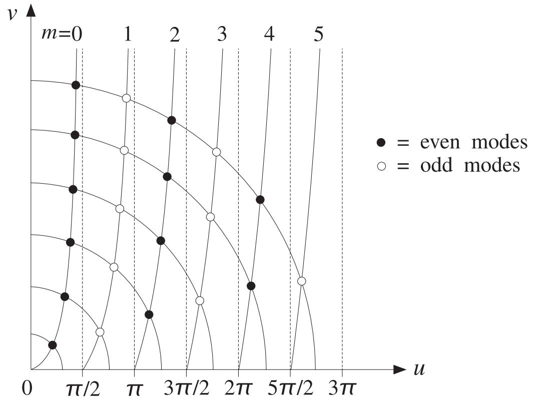


Fig. 9.11.2 Even and odd TE modes at different frequencies.

Given a value of R, we determine M as that integer satisfying Eq. (9.11.24), or, $M\le 2R/\pi<M+1$, that is, the largest integer less than $2R/\pi$:

$$\boxed{M=\text{floor}\Big(\frac{2R}{\pi}\Big)}\qquad\text{(maximum mode number)}\qquad(9.11.25)$$

Then, there will be $M+1$ solutions indexed by $m=0,1,\dots,M$, which will correspond to even modes if m is even and to odd modes if m is odd. The $M+1$ branches of $\tan u$ and $\cot u$ being intersected by the R-circle are those contained in the u-ranges:

$$\boxed{R_m\le u<R_{m+1}}\,,\quad m=0,1,\dots,M\qquad(9.11.26)$$

where

$$\boxed{R_m=\frac{m\pi}{2}}\,,\quad m=0,1,\dots,M\qquad(9.11.27)$$

If m is even, the u-range (9.11.26) defines a branch of $\tan u$, and if m is odd, a branch of $\cot u$. We can combine the even and odd cases of Eq. (9.11.22) into a single case by noting the identity:

$$\tan(u-R_m)=\begin{cases}\tan u\,, & \text{if } m \text{ is even}\\ -\cot u\,, & \text{if } m \text{ is odd}\end{cases}\qquad(9.11.28)$$

This follows from the trigonometric identity:

$$\tan(u-m\pi/2)=\frac{\sin u\cos(m\pi/2)-\cos u\sin(m\pi/2)}{\cos u\cos(m\pi/2)+\sin u\sin(m\pi/2)}$$

Therefore, to find the mth mode, whether even or odd, we must find the unique solution of the following system in the u-range $R_m\le u<R_{m+1}$:

$$\boxed{\begin{aligned} v &= u\tan(u-R_m) \\ v^2+u^2 &= R^2 \end{aligned} \qquad (m\text{th mode})} \tag{9.11.29}$$

If one had an approximate solution u, v for the mth mode, one could refine it by using Newton's method, which converges very fast provided it is close to the true solution. Just such an approximate solution, accurate to within one percent of the true solution, was given by Lotspeich [857]. Without going into the detailed justification of this method, the approximation is as follows:

$$u = R_m + w_1(m)u_1(m) + w_2(m)u_2(m)\,, \quad m = 0, 1, \dots, M \tag{9.11.30}$$

where $u_1(m)$, $u_2(m)$ are approximate solutions near and far from the cutoff R_m, and $w_1(m)$, $w_2(m)$ are weighting factors:

$$\begin{aligned} u_1(m) &= \frac{\sqrt{1+2R(R-R_m)}-1}{R}\,, \quad u_2(m) = \frac{\pi}{2}\,\frac{R-m}{R+1} \\ w_1(m) &= \exp\bigl(-(R-R_m)^2/V_m^2\bigr)\,, \quad w_2(m) = 1 - w_1(m) \\ V_m &= \frac{1}{\sqrt{\ln 1.25}}\left(\frac{\pi/4+R_m}{\cos(\pi/4)} - R_m\right) \end{aligned} \tag{9.11.31}$$

This solution serves as the starting point to Newton's iteration for solving the equation $F(u) = 0$, where $F(u)$ is defined by

$$F(u) = u\tan(u-R_m) - v = u\tan(u-R_m) - \sqrt{R^2-u^2} \tag{9.11.32}$$

Newton's iteration is:

$$\boxed{\begin{aligned} &\textit{for } i = 1, 2\dots, N_{\text{it}} \textit{ do:} \\ &\qquad u = u - \frac{F(u)}{G(u)} \end{aligned}} \tag{9.11.33}$$

where $G(u)$ is the derivative $F'(u)$, correct to order $O(F)$:

$$G(u) = \frac{v}{u} + \frac{u}{v} + \frac{R^2}{u} \tag{9.11.34}$$

The solution steps defined in Eqs. (9.11.29)–(9.11.34) have been implemented in the MATLAB function `dslab.m`, with usage:

```
[u,v,err] = dslab(R,Nit);                 % TE-mode cutoff wavenumbers in a dielectric slab
```

where N_{it} is the desired number of Newton iterations (9.11.33), `err` is the value of $F(u)$ at the end of the iterations, and u, v are the $(M+1)$-dimensional vectors of solutions. The number of iterations is typically very small, $N_{\text{it}} = 2$–3.

The related MATLAB function `dguide.m` uses `dslab` to calculate the solution parameters β, k_c, α_c, given the frequency f, the half-length a, and the refractive indices n_1, n_2 of the slab. It has usage:

```
[be,kc,ac,fc,err] = dguide(f,a,n1,n2,Nit);        % dielectric slab guide
```

where f is in GHz, a in cm, and β, k_c, α_c in cm^{-1}. The quantity f_c is the vector of the $M+1$ cutoff frequencies defined by the branch edges $R_m = m\pi/2$, that is, $R_m = \omega_m a N_A/c_0 = 2\pi f_m a N_A/c_0 = m\pi/2$, or,

$$\boxed{f_m = \frac{mc_0}{4aN_A}}\,, \quad m = 0, 1, \dots, M \tag{9.11.35}$$

The meaning of f_m is that there are $m+1$ propagating modes for each f is in the interval $f_m \le f < f_{m+1}$.

Example 9.11.1: *Dielectric Slab Waveguide.* Determine the propagating TE modes of a dielectric slab of half-length $a = 0.5$ cm at frequency $f = 30$ GHz. The refractive indices of the slab and the surrounding dielectric are $n_1 = 2$ and $n_2 = 1$.

Solution: The solution is obtained by the MATLAB call:

```
f = 30; a = 0.5; n1 = 2; n2 = 1; Nit = 3;
[be,kc,ac,fc,err] = dguide(f,a,n1,n2,Nit)
```

The frequency radius is $R = 5.4414$, which gives $2R/\pi = 3.4641$, and therefore, $M = 3$. The resulting solutions, depicted in Fig. 9.11.3, are as follows:

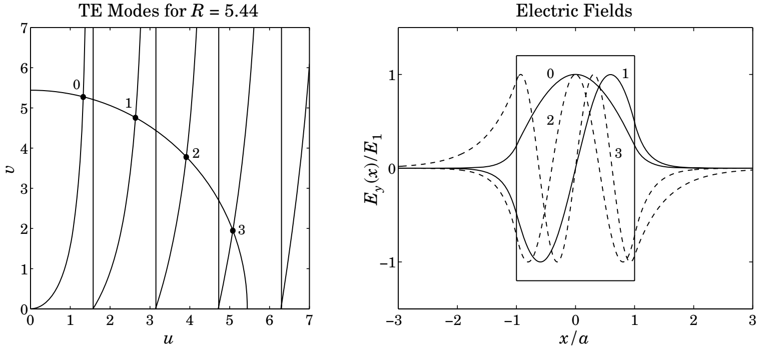


Fig. 9.11.3 TE modes and corresponding E-field patterns.

m	u	v	β	k_c	α_c	f_m
0	1.3248	5.2777	12.2838	2.6497	10.5553	0.0000
1	2.6359	4.7603	11.4071	5.2718	9.5207	8.6603
2	3.9105	3.7837	9.8359	7.8210	7.5675	17.3205
3	5.0793	1.9519	7.3971	10.1585	3.9037	25.9808

The cutoff frequencies f_m are in GHz. We note that as the mode number m increases, the quantity α_c decreases and the effective skin depth $1/\alpha_c$ increases, causing the fields outside the slab to be less confined. The electric field patterns are also shown in the figure as functions of x.

The approximation error, `err`, is found to be 4.885×10^{-15} using only three Newton iterations. Using two, one, and no (the Lotspeich approximation) iterations would result in the errors 2.381×10^{-8}, 4.029×10^{-4}, and 0.058.

The lowest non-zero cutoff frequency is $f_1 = 8.6603$ GHz, implying that there will be a single solution if f is in the interval $0 \le f < f_1$. For example, if $f = 5$ GHz, the solution is $\beta = 1.5649$ rad/cm, $k_c = 1.3920$ rad/cm, and $\alpha_c = 1.1629$ nepers/cm.

The frequency range over which there are only four solutions is $[25.9808, 34.6410]$ GHz, where the upper limit is $4f_1$. □

In terms of the ray picture of the propagating wave, the angles of total internal reflection are quantized according to the values of the propagation wavenumber β for the various modes.

If we denote by $k_1 = k_0 n_1$ the wavenumber within the slab, then the wavenumbers β, k_c are the z- and x-components k_z, k_x of k_1 with an angle of incidence θ. (The vectorial relationships are the same as those in Fig. 9.9.1.) Thus, we have:

$$\begin{aligned} \beta &= k_1 \sin\theta = k_0 n_1 \sin\theta \\ k_c &= k_1 \cos\theta = k_0 n_1 \cos\theta \end{aligned} \tag{9.11.36}$$

The value of β for each mode will generate a corresponding value for θ. The attenuation wavenumber α_c outside the slab can also be expressed in terms of the total internal reflection angles:

$$\alpha_c = \sqrt{\beta^2 - k_0^2 n_2^2} = k_0\sqrt{n_1^2\sin^2\theta - n_2^2}$$

Since the critical angle is $\sin\theta_c = n_2/n_1$, we may also express α_c as:

$$\alpha_c = k_0 n_1 \sqrt{\sin^2\theta - \sin\theta_c^2} \tag{9.11.37}$$

Example 9.11.2: For the Example 9.11.1, we calculate $k_0 = 6.2832$ and $k_1 = 12.5664$ rad/cm. The critical and total internal reflection angles of the four modes are found to be:

$$\theta_c = \text{asin}\left(\frac{n_2}{n_1}\right) = 30^{\text{o}}$$

$$\theta = \text{asin}\left(\frac{\beta}{k_1}\right) = \{77.8275^{\text{o}}, 65.1960^{\text{o}}, 51.5100^{\text{o}}, 36.0609^{\text{o}}\}$$

As required, all θs are greater than θ_c. □

9.12 Problems

9.1 An air-filled 1.5 cm×3 cm waveguide is operated at a frequency that lies in the middle of its TE_{10} mode band. Determine this operating frequency in GHz and calculate the maximum power in Watts that can be transmitted without causing dielectric breakdown of air. The dielectric strength of air is 3 MV/m.

9.2 It is desired to design an air-filled rectangular waveguide such that (a) it operates only in the TE_{10} mode with the widest possible bandwidth, (b) it can transmit the maximum possible power, and (c) the operating frequency is 12 GHz and it lies in the middle of the operating band. What are the dimensions of the guide in cm?

9.3 An air-filled rectangular waveguide is used to transfer power to a radar antenna. The guide must meet the following specifications: The two lowest modes are TE_{10} and TE_{20}. The operating frequency is 3 GHz and must lie exactly halfway between the cutoff frequencies of these two modes. The maximum electric field within the guide may not exceed, by a safety margin of 3, the breakdown field of air 3 MV/m.

a. Determine the smallest dimensions a, b for such a waveguide, if the transmitted power is required to be 1 MW.

b. What are the dimensions a, b if the transmitted power is required to be maximum? What is that maximum power in MW?

9.4 It is desired to design an air-filled rectangular waveguide operating at 5 GHz, whose group velocity is $0.8c$. What are the dimensions a, b of the guide (in cm) if it is also required to carry maximum power and have the widest possible bandwidth? What is the cutoff frequency of the guide in GHz and the operating bandwidth?

9.5 Show the following relationship between guide wavelength and group velocity in an arbitrary air-filled waveguide: $v_g\lambda_g = c\lambda$, where $\lambda_g = 2\pi/\beta$ and λ is the free-space wavelength. Moreover, show that the λ and λ_g are related to the cutoff wavelength λ_c by:

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_g^2} + \frac{1}{\lambda_c^2}$$

9.6 Determine the four lowest modes that can propagate in a WR-159 and a WR-90 waveguide. Calculate the cutoff frequencies (in GHz) and cutoff wavelengths (in cm) of these modes.

9.7 An air-filled WR-90 waveguide is operated at 9 GHz. Calculate the maximum power that can be transmitted without causing dielectric breakdown of air. Calculate the attenuation constant in dB/m due to wall ohmic losses. Assume copper walls.

9.8 A rectangular waveguide has sides a, b such that $b \le a/2$. Determine the cutoff wavelength λ_c of this guide. Show that the operating wavelength band of the lowest mode is $0.5\lambda_c \le \lambda \le \lambda_c$. Moreover, show that the allowed range of the guide wavelength is $\lambda_g \ge \lambda_c/\sqrt{3}$.

9.9 The TE_{10} mode operating bandwidth of an air-filled waveguide is required to be 4–7 GHz. What are the dimensions of the guide?

9.10 *Computer Experiment: WR-159 Waveguide.* Reproduce the two graphs of Fig. 9.8.2.

9.11 *Computer Experiment: Dielectric Slab Waveguide.* Using the MATLAB functions `dslab` and `dguide`, write a program that reproduces all the results and graphs of Examples 9.11.1 and 9.11.2.

9.12 A TM mode is propagated along a waveguide of arbitrary but uniform cross section. Assume perfectly conducting walls.

a. Show that the $E_z(x, y)$ component satisfies:

$$\int_S |\nabla E_z|^2\, dS = k_c^2 \int_S |E_z|^2\, dS$$

b. Using the above result, show that the energy velocity is equal to the group velocity.

10

Transmission Lines

10.1 General Properties of TEM Transmission Lines

We saw in Sec. 9.3 that TEM modes are described by Eqs. (9.3.3) and (9.3.4), the latter being equivalent to a two-dimensional electrostatic problem:

$$\boxed{\begin{aligned} \boldsymbol{H}_T &= \frac{1}{\eta}\,\hat{\mathbf{z}}\times\boldsymbol{E}_T \\ \boldsymbol{\nabla}_T\times\boldsymbol{E}_T &= 0 \\ \boldsymbol{\nabla}_T\cdot\boldsymbol{E}_T &= 0 \end{aligned}} \qquad \text{(TEM modes)} \tag{10.1.1}$$

The second of (10.1.1) implies that $\boldsymbol{E}_T$ can be expressed as the (two-dimensional) gradient of a scalar electrostatic potential. Then, the third equation becomes Laplace's equation for the potential. Thus, the electric field can be obtained from:

$$\boxed{\begin{aligned} \nabla_T^2\varphi &= 0 \\ \boldsymbol{E}_T &= -\boldsymbol{\nabla}_T\varphi \end{aligned}} \qquad \text{(equivalent electrostatic problem)} \tag{10.1.2}$$

Because in electrostatic problems the electric field lines must start at positively charged conductors and end at negatively charged ones, a TEM mode can be supported only in multi-conductor guides, such as the coaxial cable or the two-wire line. Hollow conducting waveguides cannot support TEM modes.

Fig. 10.1.1 depicts the transverse cross-sectional area of a two-conductor transmission line. The cross-section shapes are arbitrary.

The conductors are *equipotentials* of the electrostatic solution. Let φ_a, φ_b be the constant potentials on the two conductors. The voltage difference between the conductors will be $V = \varphi_a - \varphi_b$. The electric field lines start perpendicularly on conductor (a) and end perpendicularly on conductor (b).

The magnetic field lines, being perpendicular to the electric lines according to Eq. (10.1.1), are recognized to be the *equipotential lines*. As such, they close upon themselves surrounding the two conductors.

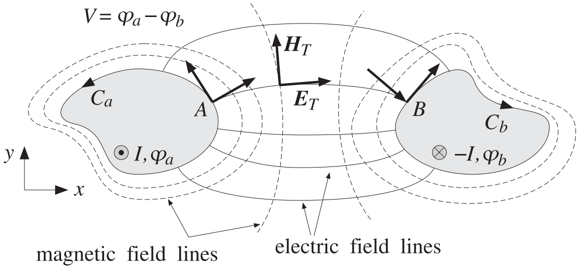


Fig. 10.1.1 Two-conductor transmission line.

In particular, on the conductor surfaces the magnetic field is tangential. According to Ampère's law, the line integrals of the magnetic field around each conductor will result into total currents I and $-I$ flowing on the conductors in the z-direction. These currents are equal and opposite.

Impedance, Inductance, and Capacitance

Because the fields are propagating along the z-direction with frequency ω and wavenumber $\beta=\omega/c$, the z,t dependence of the voltage V and current I will be:

$$\boxed{\begin{aligned} V(z,t) &= Ve^{j\omega t-j\beta z}\\ I(z,t) &= Ie^{j\omega t-j\beta z}\end{aligned}} \tag{10.1.3}$$

For backward-moving voltage and current waves, we must replace β by $-\beta$. The ratio $V(z,t)/I(z,t)=V/I$ remains constant and independent of z. It is called the *characteristic impedance* of the line:

$$\boxed{Z=\frac{V}{I}} \qquad \text{(line impedance)} \tag{10.1.4}$$

In addition to the impedance Z, a TEM line is characterized by its inductance per unit length L' and its capacitance per unit length C'. For lossless lines, the three quantities Z, L', C' are related as follows:

$$\boxed{L'=\mu\frac{Z}{\eta}\,,\quad C'=\epsilon\frac{\eta}{Z}} \qquad \text{(inductance and capacitance per unit length)} \tag{10.1.5}$$

where $\eta=\sqrt{\mu/\epsilon}$ is the characteristic impedance of the dielectric medium between the conductors.[†] By multiplying and dividing L' and C', we also obtain:

$$\boxed{Z=\sqrt{\frac{L'}{C'}}\,,\quad c=\frac{1}{\sqrt{\epsilon\mu}}=\frac{1}{\sqrt{L'C'}}} \tag{10.1.6}$$

[†] These expressions explain why μ and ϵ are sometimes given in units of henry/m and farad/m.

The *velocity factor* of the line is the ratio $c/c_0=1/n$, where $n=\sqrt{\epsilon/\epsilon_0}=\sqrt{\epsilon_r}$ is the refractive index of the dielectric, which is assumed to be non-magnetic.

Because $\omega=\beta c$, the guide wavelength will be $\lambda=2\pi/\beta=c/f=c_0/fn=\lambda_0/n$, where λ_0 is the free-space wavelength. For a finite length l of the transmission line, the quantity $l/\lambda=nl/\lambda_0$ is referred to as the *electrical length* of the line and plays the same role as the optical length in thin-film layers.

Eqs. (10.1.5) and (10.1.6) are general results that are valid for any TEM line. They can be derived with the help of Fig. 10.1.2.

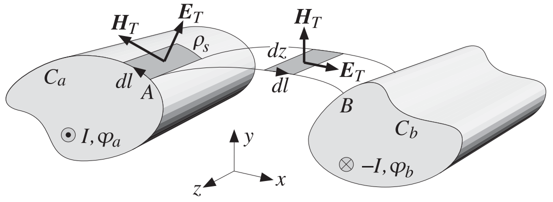


Fig. 10.1.2 Surface charge and magnetic flux linkage.

The voltage V is obtained by integrating $\mathbf{E}_T\cdot d\mathbf{l}$ along any path from (a) to (b). However, if that path is chosen to be an E-field line, then $\mathbf{E}_T\cdot d\mathbf{l}=|\mathbf{E}_T|dl$, giving:

$$V=\int_a^b|\mathbf{E}_T|dl \tag{10.1.7}$$

Similarly, the current I can be obtained by the integral of $\mathbf{H}_T\cdot d\mathbf{l}$ along any closed path around conductor (a). If that path is chosen to be an H-field line, such as the periphery C_a of the conductor, we will obtain:

$$I=\oint_{C_a}|\mathbf{H}_T|dl \tag{10.1.8}$$

The surface charge accumulated on an infinitesimal area $dl\,dz$ of conductor (a) is $dQ=\rho_s dl\,dz$, where ρ_s is the surface charge density. Because the conductors are assumed to be perfect, the boundary conditions require that ρ_s be equal to the normal component of the D-field, that is, $\rho_s=\epsilon|\mathbf{E}_T|$. Thus, $dQ=\epsilon|\mathbf{E}_T|dl\,dz$.

If we integrate over the periphery C_a of conductor (a), we will obtain the total surface charge per unit z-length:

$$Q'=\frac{dQ}{dz}=\oint_{C_a}\epsilon|\mathbf{E}_T|dl$$

But because of the relationship $|\mathbf{E}_T|=\eta|\mathbf{H}_T|$, which follows from the first of Eqs. (10.1.1), we have:

$$Q'=\oint_{C_a}\epsilon|\mathbf{E}_T|dl=\epsilon\eta\oint_{C_a}|\mathbf{H}_T|dl=\epsilon\eta I \tag{10.1.9}$$

where we used Eq. (10.1.8). Because Q' is related to the capacitance per unit length and the voltage by $Q'=C'V$, we obtain

$$\boxed{Q' = C'V = \epsilon\eta I} \quad\Rightarrow\quad C' = \epsilon\eta\,\frac{I}{V} = \epsilon\,\frac{\eta}{Z}$$

Next, we consider an E-field line between points A and B on the two conductors. The magnetic flux through the infinitesimal area $dl\,dz$ will be $d\Phi = |\mathbf{B}_T|dl\,dz = \mu|\mathbf{H}_T|dl\,dz$ because the vector $\mathbf{H}_T$ is perpendicular to the area.

If we integrate from (a) to (b), we will obtain the total magnetic flux linking the two conductors per unit z-length:

$$\Phi' = \frac{d\Phi}{dz} = \int_a^b \mu|\mathbf{H}_T|dl$$

replacing $|\mathbf{H}_T| = |\mathbf{E}_T|/\eta$ and using Eq. (10.1.7), we find:

$$\Phi' = \int_a^b \mu|\mathbf{H}_T|dl = \frac{\mu}{\eta}\int_a^b |\mathbf{E}_T|dl = \frac{\mu}{\eta}V$$

The magnetic flux is related to the inductance via $\Phi' = L'I$. Therefore, we get:

$$\boxed{\Phi' = L'I = \frac{\mu}{\eta}V} \quad\Rightarrow\quad L' = \frac{\mu}{\eta}\frac{V}{I} = \mu\,\frac{Z}{\eta}$$

Transmitted Power

The relationships among Z, L', C' can also be derived using energy considerations. The power transmitted along the line is obtained by integrating the z-component of the Poynting vector over the cross-section S of the line. For TEM modes we have $\mathcal{P}_z = |\mathbf{E}_T|^2/2\eta$, therefore,

$$P_T = \frac{1}{2\eta}\iint_S |\mathbf{E}_T|^2 dx\,dy = \frac{1}{2\eta}\iint_S |\nabla_T\varphi|^2 dx\,dy \tag{10.1.10}$$

It can be shown in general that Eq. (10.1.10) can be rewritten as:

$$\boxed{P_T = \frac{1}{2}\,\mathrm{Re}(V^*I) = \frac{1}{2}Z|I|^2 = \frac{1}{2Z}|V|^2} \tag{10.1.11}$$

We will verify this in the various examples below. It can be proved using the following Green's identity:

$$|\nabla_T\varphi|^2 + \varphi^*\nabla_T^2\varphi = \nabla_T\cdot(\varphi^*\nabla_T\varphi)$$

Writing $\mathbf{E}_T = -\nabla_T\varphi$ and noting that $\nabla_T^2\varphi = 0$, we obtain:

$$|\mathbf{E}_T|^2 = -\nabla_T\cdot(\varphi^*\mathbf{E}_T)$$

Then, the two-dimensional Gauss' theorem implies:

$$\begin{aligned}
P_T &= \frac{1}{2\eta}\iint_S |\mathbf{E}_T|^2 dx\,dy = -\frac{1}{2\eta}\iint_S \nabla_T\cdot(\varphi^*\mathbf{E}_T)\,dx\,dy \\
&= -\frac{1}{2\eta}\oint_{C_a}\varphi^*\mathbf{E}_T\cdot(-\hat{\mathbf{n}})\,dl - \frac{1}{2\eta}\oint_{C_b}\varphi^*\mathbf{E}_T\cdot(-\hat{\mathbf{n}})\,dl \\
&= \frac{1}{2\eta}\oint_{C_a}\varphi^*\,(\mathbf{E}_T\cdot\hat{\mathbf{n}})\,dl + \frac{1}{2\eta}\oint_{C_b}\varphi^*\,(\mathbf{E}_T\cdot\hat{\mathbf{n}})\,dl
\end{aligned}$$

where $\hat{\mathbf{n}}$ are the outward normals to the conductors (the quantity $-\hat{\mathbf{n}}$ is the normal outward from the region S.) Because the conductors are equipotential surfaces, we have $\varphi^* = \varphi_a^*$ on conductor (a) and $\varphi^* = \varphi_b^*$ on conductor (b). Using Eq. (10.1.9) and noting that $\mathbf{E}_T\cdot\hat{\mathbf{n}} = \pm|\mathbf{E}_T|$ on conductors (a) and (b), we obtain:

$$\begin{aligned}
P_T &= \frac{1}{2\eta}\varphi_a^*\oint_{C_a}|\mathbf{E}_T|dl - \frac{1}{2\eta}\varphi_b^*\oint_{C_b}|\mathbf{E}_T|dl = \frac{1}{2\eta}\varphi_a^*\frac{Q'}{\epsilon} - \frac{1}{2\eta}\varphi_b^*\frac{Q'}{\epsilon} \\
&= \frac{1}{2}(\varphi_a^* - \varphi_b^*)\frac{Q'}{\epsilon\eta} = \frac{1}{2}V^*\frac{\epsilon\eta I}{\epsilon\eta} = \frac{1}{2}V^*I = \frac{1}{2}Z|I|^2
\end{aligned}$$

The distribution of electromagnetic energy along the line is described by the time-averaged electric and magnetic energy densities per unit length, which are given by:

$$W_e' = \frac{1}{4}\epsilon\iint_S |\mathbf{E}_T|^2 dx\,dy\,, \quad W_m' = \frac{1}{4}\mu\iint_S |\mathbf{H}_T|^2 dx\,dy$$

Using Eq. (10.1.10), we may rewrite:

$$W_e' = \frac{1}{2}\epsilon\eta P_T = \frac{1}{2c}P_T\,, \quad W_m' = \frac{1}{2}\frac{\mu}{\eta}P_T = \frac{1}{2c}P_T$$

Thus, $W_e' = W_m'$ and the total energy density is $W' = W_e' + W_m' = P_T/c$, which implies that the energy velocity will be $v_{\text{en}} = P_T/W' = c$. We may also express the energy densities in terms of the capacitance and inductance of the line:

$$W_e' = \frac{1}{4}C'|V|^2\,, \quad W_m' = \frac{1}{4}L'|I|^2 \tag{10.1.12}$$

Power Losses, Resistance, and Conductance

Transmission line losses can be handled in the manner discussed in Sec. 9.2. The field patterns and characteristic impedance are determined assuming the conductors are perfectly conducting. Then, the losses due to the ohmic heating of the dielectric and the conductors can be calculated by Eqs. (9.2.5) and (9.2.9).

These losses can be quantified by two more characteristic parameters of the line, the resistance and conductance per unit length, R' and G'. The attenuation coefficients due to conductor and dielectric losses are then expressible in terms R', G' and Z by:

$$\boxed{\alpha_c = \frac{R'}{2Z}\,, \quad \alpha_d = \frac{1}{2}G'Z} \tag{10.1.13}$$

They can be derived in general terms as follows. The induced surface currents on the conductor walls are $\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H}_T = \hat{\mathbf{n}}\times(\hat{\mathbf{z}}\times\boldsymbol{E}_T)/\eta$, where $\hat{\mathbf{n}}$ is the outward normal to the wall.

Using the BAC-CAB rule, we find $\boldsymbol{J}_s = \hat{\mathbf{z}}(\hat{\mathbf{n}}\cdot\boldsymbol{E}_T)/\eta$. But, $\hat{\mathbf{n}}$ is parallel to $\boldsymbol{E}_T$ on the surface of conductor (a), and anti parallel on (b). Therefore, $\hat{\mathbf{n}}\cdot\boldsymbol{E}_T = \pm|\boldsymbol{E}_T|$. It follows that $\boldsymbol{J}_s = \pm\hat{\mathbf{z}}|\boldsymbol{E}_T|/\eta = \pm\hat{\mathbf{z}}|\boldsymbol{H}_T|$, pointing in the $+z$ direction on (a) and $-z$ direction on (b). Inserting these expressions into Eq. (9.2.8), we find for the conductor power loss per unit z-length:

$$\boxed{P'_{\text{loss}} = \frac{dP_{\text{loss}}}{dz} = \frac{1}{2}R_s\oint_{C_a}|\boldsymbol{H}_T|^2\,dl + \frac{1}{2}R_s\oint_{C_b}|\boldsymbol{H}_T|^2\,dl} \tag{10.1.14}$$

Because $\boldsymbol{H}_T$ is related to the total current I via Eq. (10.1.8), we may define the resistance per unit length R' through the relationship:

$$\boxed{P'_{\text{loss}} = \frac{1}{2}R'|I|^2} \qquad \text{(conductor ohmic losses)} \tag{10.1.15}$$

Using Eq. (10.1.11), we find for the attenuation coefficient:

$$\alpha_c = \frac{P'_{\text{loss}}}{2P_T} = \frac{\frac{1}{2}R'|I|^2}{2\frac{1}{2}Z|I|^2} = \frac{R'}{2Z} \tag{10.1.16}$$

If the dielectric between the conductors is slightly conducting with conductivity σ_d or loss tangent $\tan\delta = \sigma_d/\epsilon\omega$, then there will be some current flow between the two conductors.

The induced shunt current per unit z-length is related to the conductance by $I'_d = G'V$. The shunt current density within the dielectric is $\boldsymbol{J}_d = \sigma_d\boldsymbol{E}_T$. The total shunt current flowing out of conductor (a) towards conductor (b) is obtained by integrating $\boldsymbol{J}_d$ around the periphery of conductor (a):

$$I'_d = \oint_{C_a}\boldsymbol{J}_d\cdot\hat{\mathbf{n}}\,dl = \sigma_d\oint_{C_a}|\boldsymbol{E}_T|dl$$

Using Eq. (10.1.9), we find:

$$I'_d = \sigma_d\frac{Q'}{\epsilon} = G'V \quad\Rightarrow\quad \boxed{G' = \frac{\sigma_d}{\epsilon}C' = \sigma_d\frac{\eta}{Z}}$$

It follows that the dielectric loss constant (9.2.5) will be:

$$\alpha_d = \frac{1}{2}\sigma_d\eta = \frac{1}{2}G'Z$$

Alternatively, the power loss per unit length due to the shunt current will be $P'_d = \text{Re}(I'_dV^*)/2 = G'|V|^2/2$, and therefore, α_d can be computed from:

$$\alpha_d = \frac{P'_d}{2P_T} = \frac{\frac{1}{2}G'|V|^2}{2\frac{1}{2Z}|V|^2} = \frac{1}{2}G'Z$$

It is common practice to express the dielectric losses and shunt conductance in terms of the loss tangent $\tan\delta$ and the wavenumber $\beta = \omega/c = \omega\epsilon\eta$:

$$\boxed{\alpha_d = \frac{1}{2}\sigma_d\eta = \frac{1}{2}\omega\epsilon\eta\tan\delta = \frac{1}{2}\beta\tan\delta} \quad\text{and}\quad \boxed{G' = \frac{\sigma_d}{\epsilon}C' = \omega C'\tan\delta} \tag{10.1.17}$$

Next, we discuss four examples: the parallel plate line, the microstrip line, the coaxial cable, and the two-wire line. In each case, we discuss the nature of the electrostatic problem and determine the characteristic impedance Z and the attenuation coefficients α_c and α_d.

10.2 Parallel Plate Lines

The parallel plate line shown in Fig. 10.2.1 consists of two parallel conducting plates of width w separated by height h by a dielectric material ϵ. Examples of such lines are microstrip lines used in microwave integrated circuits.

For arbitrary values of w and h, the fringing effects at the ends of the plates cannot be ignored. In fact, fringing requires the fields to have longitudinal components, and therefore TEM modes are not strictly-speaking supported.

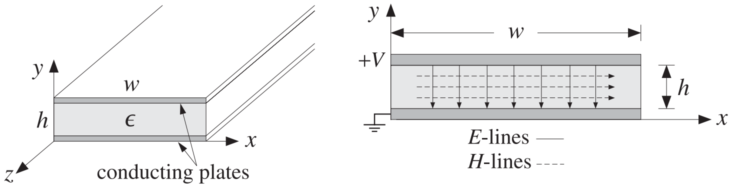


Fig. 10.2.1 Parallel plate transmission line.

However, assuming the width is much larger than the height, $w \gg h$, we may ignore the fringing effects and assume that the fields have no dependence on the x-coordinate.

The electrostatic problem is equivalent to that of a parallel plate capacitor. Thus, the electric field will have only a y component and will be constant between the plates. Similarly, the magnetic field will have only an x component. It follows from Eqs. (10.1.7) and (10.1.8) that:

$$V = -E_y h\,, \quad I = H_x w$$

Therefore, the characteristic impedance of the line will be:

$$Z = \frac{V}{I} = \frac{-E_y h}{H_x w} = \eta\frac{h}{w} \tag{10.2.1}$$

where we used $E_y = -\eta H_x$. The transmitted power is obtained from Eq. (10.1.10):

$$P_T = \frac{1}{2\eta}|E_y|^2(wh) = \frac{1}{2\eta}\frac{V^2}{h^2}wh = \frac{1}{2\eta}\frac{w}{h}V^2 = \frac{1}{2Z}V^2 = \frac{1}{2}ZI^2 \tag{10.2.2}$$

The inductance and capacitance per unit length are obtained from Eq. (10.1.5):

$$L' = \mu\frac{h}{w}\,,\quad C' = \epsilon\frac{w}{h} \tag{10.2.3}$$

The surface current on the top conductor is $\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H} = (-\hat{\mathbf{y}})\times\boldsymbol{H} = \hat{\mathbf{z}}\,H_x$. On the bottom conductor, it will be $\boldsymbol{J}_s = -\hat{\mathbf{z}}H_x$. Therefore, the power loss per unit z-length is obtained from Eq. (9.2.8):

$$P'_{\text{loss}} = 2\frac{1}{2}R_s|H_x|^2 w = \frac{1}{w}R_s I^2$$

Comparing with Eq. (10.1.15), we identify the resistance per unit length $R' = 2R_s/w$. Then, the attenuation constant due to conductor losses will be:

$$\alpha_c = \frac{P'_{\text{loss}}}{2P_T} = \frac{R'}{2Z} = \frac{R_s}{wZ} = \frac{R_s}{h\eta} \tag{10.2.4}$$

10.3 Microstrip Lines

Practical microstrip lines, shown in Fig. 10.3.1, have width-to-height ratios w/h that are not necessarily much greater than unity, and can vary over the interval $0.1 < w/h < 10$. Typical heights h are of the order of millimeters.

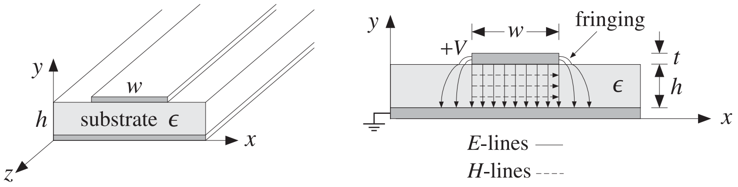


Fig. 10.3.1 A microstrip transmission line.

Fringing effects cannot be ignored completely and the simple assumptions about the fields of the parallel plate line are not valid. For example, assuming a propagating wave in the z-direction with z, t dependence of $e^{j\omega t - j\beta z}$ with a common β in the dielectric and air, the longitudinal-transverse decomposition (9.1.5) gives:

$$\nabla_T E_z\times\hat{\mathbf{z}} - j\beta\,\hat{\mathbf{z}}\times\boldsymbol{E}_T = -j\omega\mu\boldsymbol{H}_T \quad\Rightarrow\quad \hat{\mathbf{z}}\times(\nabla_T E_z + j\beta\boldsymbol{E}_T) = j\omega\mu\boldsymbol{H}_T$$

In particular, we have for the x-component:

$$\partial_y E_z + j\beta E_y = -j\omega\mu H_x$$

The boundary conditions require that the components H_x and $D_y = \epsilon E_y$ be continuous across the dielectric-air interface (at $y = h$). This gives the interface conditions:

$$\begin{aligned}\partial_y E_z^{\text{air}} + j\beta E_y^{\text{air}} &= \partial_y E_z^{\text{diel}} + j\beta E_y^{\text{diel}}\\ \epsilon_0 E_y^{\text{air}} &= \epsilon E_y^{\text{diel}}\end{aligned}$$

Combining the two conditions, we obtain:

$$\partial_y\big(E_z^{\text{diel}} - E_z^{\text{air}}\big) = j\beta\frac{\epsilon-\epsilon_0}{\epsilon}E_y^{\text{air}} = j\beta\frac{\epsilon-\epsilon_0}{\epsilon_0}E_y^{\text{diel}} \tag{10.3.1}$$

Because E_y is non-zero on either side of the interface, it follows that the left-hand side of Eq. (10.3.1) cannot be zero and the wave cannot be assumed to be strictly TEM.

However, E_y is small in both the air and the dielectric in the fringing regions (to the left and right of the upper conductor). This gives rise to the so-called *quasi-TEM approximation* in which the fields are assumed to be approximately TEM and the effect of the deviation from TEM is taken into account by empirical formulas for the line impedance and velocity factor.

In particular, the air-dielectric interface is replaced by an *effective* dielectric, filling uniformly the entire space, and in which there would be a TEM propagating mode. If we denote by ϵ_{eff} the relative permittivity of the effective dielectric, the wavelength and velocity factor of the line will be given in terms of their free-space values λ_0, c_0:

$$\boxed{\lambda = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}}\,,\quad c = \frac{c_0}{\sqrt{\epsilon_{\text{eff}}}}} \tag{10.3.2}$$

There exist many empirical formulas for the characteristic impedance of the line and the effective dielectric constant. Hammerstad and Jensen's are some of the most accurate ones [862,868]:

$$\boxed{\epsilon_{\text{eff}} = \frac{\epsilon_r+1}{2} + \frac{\epsilon_r-1}{2}\left(1+\frac{10}{u}\right)^{-ab}}\,,\quad u = \frac{w}{h} \tag{10.3.3}$$

where $\epsilon_r = \epsilon/\epsilon_0$ is the relative permittivity of the dielectric and the quantities a, b are defined by:

$$\begin{aligned}a &= 1 + \frac{1}{49}\ln\left[\frac{u^4+(u/52)^2}{u^4+0.432}\right] + \frac{1}{18.7}\ln\left[1+\left(\frac{u}{18.1}\right)^3\right]\\ b &= 0.564\left(\frac{\epsilon_r-0.9}{\epsilon_r+3}\right)^{0.053}\end{aligned} \tag{10.3.4}$$

The accuracy of these formulas is better than 0.01% for $u < 1$ and 0.03% for $u < 1000$. Similarly, the characteristic impedance is given by the empirical formula:

$$\boxed{Z = \frac{\eta_0}{2\pi\sqrt{\epsilon_{\text{eff}}}}\ln\left[\frac{f(u)}{u} + \sqrt{1+\frac{4}{u^2}}\,\right]} \tag{10.3.5}$$

where $\eta_0 = \sqrt{\mu_0/\epsilon_0}$ and the function $f(u)$ is defined by:

$$f(u) = 6 + (2\pi-6)\exp\left[-\left(\frac{30.666}{u}\right)^{0.7528}\right] \tag{10.3.6}$$

The accuracy is better than 0.2% for $0.1 \le u \le 100$ and $\epsilon_r < 128$. In the limit of large ratio w/h, or, $u\to\infty$, Eqs. (10.3.3) and (10.3.5) tend to those of the parallel plate line of the previous section:

$$\epsilon_{\text{eff}} \to \epsilon_r\,, \quad Z \to \frac{\eta_0}{\sqrt{\epsilon_r}}\frac{h}{w} = \eta\frac{h}{w}$$

Some typical substrate dielectric materials used in microstrip lines are alumina, a ceramic form of Al_2O_4 with $e_r = 9.8$, and RT-Duroid, a teflon composite material with $\epsilon_r = 2.2$. Practical values of the width-to-height ratio are in the range $0.1 \le u \le 10$ and practical values of characteristic impedances are between 10–200 ohm. Fig. 10.3.2 shows the dependence of Z and ϵ_{eff} on u for the two cases of $\epsilon_r = 2.2$ and $\epsilon_r = 9.8$.

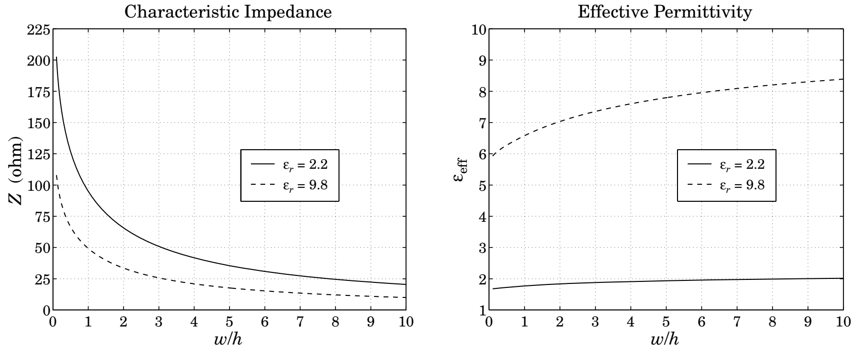


Fig. 10.3.2 Characteristic impedance and effective permittivity of microstrip line.

The synthesis of a microstrip line requires that we determine the ratio w/h that will achieve a given characteristic impedance Z. The inverse of Eq. (10.3.5)—solving for u in terms of Z—is not practical. Direct synthesis empirical equations exist [863,868], but are not as accurate as (10.3.5). Given a desired Z, the ratio $u = w/h$ is calculated as follows. If $u \le 2$,

$$u = \frac{8}{e^A - 2e^{-A}} \tag{10.3.7}$$

and, if $u > 2$,

$$u = \frac{\epsilon_r - 1}{\pi\epsilon_r}\left[\ln(B-1) + 0.39 - \frac{0.61}{\epsilon_r}\right] + \frac{2}{\pi}\big[B - 1 - \ln(2B-1)\big] \tag{10.3.8}$$

where A, B are given by:

$$\begin{aligned} A &= \pi\sqrt{2(\epsilon_r+1)}\,\frac{Z}{\eta_0} + \frac{\epsilon_r - 1}{\epsilon_r + 1}\left(0.23 + \frac{0.11}{\epsilon_r}\right) \\ B &= \frac{\pi}{2\sqrt{\epsilon_r}}\,\frac{\eta_0}{Z} \end{aligned} \tag{10.3.9}$$

The accuracy of these formulas is about 1%. The method can be improved iteratively by a process of refinement to achieve essentially the same accuracy as Eq. (10.3.5). Starting with u computed from Eqs. (10.3.7) and (10.3.8), a value of Z is computed through Eq. (10.3.5). If that Z is more than, say, 0.2% off from the desired value of the line impedance, then u is slightly changed, and so on, until the desired level of accuracy is reached [868]. Because Z is monotonically decreasing with u, if Z is less than the desired value, then u is decreased by a small percentage, else, u is increased by the same percentage.

The three MATLAB functions `mstripa`, `mstrips`, and `mstripr` implement the analysis, synthesis, and refinement procedures. They have usage:

```
[eff,Z] = mstripa(er,u);         % analysis equations (10.3.3) and (10.3.5)
u = mstrips(er,Z);               % synthesis equations (10.3.7) and (10.3.8)
[u,N] = mstripr(er,Z,per);       % refinement
```

The function `mstripa` accepts also a vector of several u's, returning the corresponding vector of values of ϵ_{eff} and Z. In `mstripr`, the output N is the number of iterations required for convergence, and `per` is the desired percentage error, which defaults to 0.2% if this parameter is omitted.

Example 10.3.1: Given $\epsilon_r = 2.2$ and $u = w/h = 2, 4, 6$, the effective permittivities and impedances are computed from the MATLAB call:

```
u = [2; 4; 6];
[eff, Z] = mstripa(er,u);
```

The resulting output vectors are:

$$u = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \quad\Rightarrow\quad \epsilon_{\text{eff}} = \begin{bmatrix} 1.8347 \\ 1.9111 \\ 1.9585 \end{bmatrix}, \quad Z = \begin{bmatrix} 65.7273 \\ 41.7537 \\ 30.8728 \end{bmatrix} \text{ohm}$$

Example 10.3.2: To compare the outputs of `mstrips` and `mstripr`, we design a microstrip line with $\epsilon_r = 2.2$ and characteristic impedance $Z = 50$ ohm. We find:

$$\begin{aligned} u &= \text{mstrips}(2.2, 50) = 3.0779 \quad\Rightarrow\quad [\epsilon_{\text{eff}}, Z] = \text{mstripa}(2.2, u) = [1.8811,\ 50.0534] \\ u &= \text{mstripr}(2.2, 50) = 3.0829 \quad\Rightarrow\quad [\epsilon_{\text{eff}}, Z] = \text{mstripa}(2.2, u) = [1.8813,\ 49.9990] \end{aligned}$$

The first solution has an error of 0.107% from the desired 50 ohm impedance, and the second, a 0.002% error.

As another example, if $Z = 100\ \Omega$, the function `mstrips` results in $u = 0.8949$, $Z = 99.9495\ \Omega$, and a 0.050% error, whereas `mstripr` gives $u = 0.8939$, $Z = 99.9980\ \Omega$, and a 0.002% error. □

In using microstrip lines several other effects must be considered, such as finite strip thickness, frequency dispersion, dielectric and conductor losses, radiation, and surface waves. Guidelines for such effects can be found in [862–868].

The dielectric losses are obtained from Eq. (10.1.17) by multiplying it by an effective dielectric *filling* factor q:

$$\alpha_d = q\frac{\omega}{2c}\tan\delta = \frac{f}{c_0}\pi q\sqrt{\epsilon_{\text{eff}}}\tan\delta = \frac{1}{\lambda_0}\pi q\sqrt{\epsilon_{\text{eff}}}\tan\delta\,, \quad q = \frac{1-\epsilon_{\text{eff}}^{-1}}{1-\epsilon_r^{-1}} \tag{10.3.10}$$

Typical values of the loss tangent are of the order of 0.001 for alumina and duroid substrates. The conductor losses are approximately computed from Eq. (10.2.4):

$$\alpha_c = \frac{R_s}{wZ} \tag{10.3.11}$$

10.4 Coaxial Lines

The coaxial cable, depicted in Fig. 10.4.1, is the most widely used TEM transmission line. It consists of two concentric conductors of inner and outer radii of a and b, with the space between them filled with a dielectric ϵ, such as polyethylene or teflon.

The equivalent electrostatic problem can be solved conveniently in cylindrical coordinates ρ, ϕ. The potential $\varphi(\rho,\phi)$ satisfies Laplace's equation:

$$\nabla_T^2\varphi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\varphi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\varphi}{\partial^2\phi} = 0$$

Because of the cylindrical symmetry, the potential does not depend on the azimuthal angle ϕ. Therefore,

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\varphi}{\partial\rho}\right) = 0 \quad\Rightarrow\quad \rho\frac{\partial\varphi}{\partial\rho} = B \quad\Rightarrow\quad \varphi(\rho) = A + B\ln\rho$$

where A, B are constants of integration. Assuming the outer conductor is grounded, $\varphi(\rho) = 0$ at $\rho = b$, and the inner conductor is held at voltage V, $\varphi(a) = V$, the constants A, B are determined to be $B = -V\ln(b/a)$ and $A = -B\ln b$, resulting in the potential:

$$\varphi(\rho) = \frac{V}{\ln(b/a)}\ln(b/\rho) \tag{10.4.1}$$

It follows that the electric field will have only a radial component, $E_\rho = -\partial_\rho\varphi$, and the magnetic field only an azimuthal component $H_\phi = E_\rho/\eta$:

$$E_\rho = \frac{V}{\ln(b/a)}\frac{1}{\rho}, \qquad H_\phi = \frac{V}{\eta\ln(b/a)}\frac{1}{\rho} \tag{10.4.2}$$

Integrating H_ϕ around the inner conductor we obtain the current:

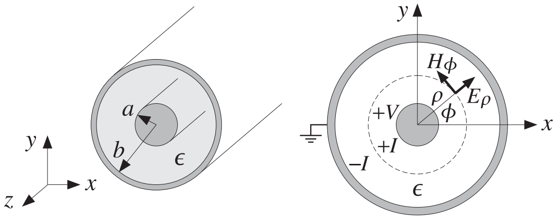


Fig. 10.4.1 Coaxial transmission line.

$$I = \int_0^{2\pi} H_\phi\,\rho\,d\phi = \int_0^{2\pi}\frac{V}{\eta\ln(b/a)}\frac{1}{\rho}\,\rho\,d\phi = \frac{2\pi V}{\eta\ln(b/a)} \tag{10.4.3}$$

It follows that the characteristic impedance of the line $Z = V/I$, and hence the inductance and capacitance per unit length, will be:

$$\boxed{Z = \frac{\eta}{2\pi}\ln(b/a), \quad L' = \frac{\mu}{2\pi}\ln(b/a), \quad C' = \frac{2\pi\epsilon}{\ln(b/a)}} \tag{10.4.4}$$

Using Eq. (10.4.3) into (10.4.2), we may express the magnetic field in the form:

$$H_\phi = \frac{I}{2\pi\rho} \tag{10.4.5}$$

This is also obtainable by the direct application of Ampère's law around the loop of radius ρ encircling the inner conductor, that is, $I = (2\pi\rho)H_\phi$.

The transmitted power can be expressed either in terms of the voltage V or in terms of the maximum value of the electric field inside the line, which occurs at $\rho = a$, that is, $E_a = V/\big(a\ln(b/a)\big)$:

$$\boxed{P_T = \frac{1}{2Z}|V|^2 = \frac{\pi|V|^2}{\eta\ln(b/a)} = \frac{1}{\eta}|E_a|^2(\pi a^2)\ln(b/a)} \tag{10.4.6}$$

Example 10.4.1: A commercially available polyethylene-filled RG-58/U cable is quoted to have impedance of 53.5 Ω, velocity factor of 66 percent, inner conductor radius $a = 0.406$ mm (AWG 20-gauge wire), and maximum operating RMS voltage of 1900 volts. Determine the outer-conductor radius b, the capacitance per unit length C', the maximum power P_T that can be transmitted, and the maximum electric field inside the cable. What should be the outer radius b if the impedance were required to be exactly 50 Ω?

Solution: Polyethylene has a relative dielectric constant of $\epsilon_r = 2.25$, so that $n = \sqrt{\epsilon_r} = 1.5$. The velocity factor is $c/c_0 = 1/n = 0.667$. Given that $\eta = \eta_0/n = 376.73/1.5 = 251.15$ Ω, we have:

$$Z = \frac{\eta}{2\pi}\ln(b/a) \quad\Rightarrow\quad b = ae^{2\pi Z/\eta} = 0.406e^{2\pi 53.5/251.15} = 1.548\ \text{mm}$$

Therefore, $b/a = 3.81$. If $Z = 50$, the above calculation would give $b = 1.418$ mm and $b/a = 3.49$. The capacitance per unit length is found from:

$$C' = \epsilon\frac{\eta}{Z} = \frac{1}{cZ} = \frac{n}{c_0 Z} = \frac{1.5}{3{\times}10^8{\times}53.5} = 93.46\ \text{pF/m}$$

For $Z = 50$ Ω, we find $C' = 100$ pF/m. The peak voltage is related to its RMS value by $|V| = \sqrt{2}V_{\text{rms}}$. It follows that the maximum power transmitted is:

$$P_T = \frac{1}{2Z}|V|^2 = \frac{V_{\text{rms}}^2}{Z} = \frac{1900^2}{53.5} = 67.5\ \text{kW}$$

The peak value of the electric field occurring at the inner conductor will be:

$$|E_a| = \frac{|V|}{a\ln(b/a)} = \frac{\sqrt{2}V_{\rm rms}}{a\ln(b/a)} = \frac{\sqrt{2}1900}{0.406\times10^{-2}\ln(3.096/0.406)} = 0.5\ \text{MV/m}$$

This is to be compared with the dielectric breakdown of air of 3 MV/m. For a 73-Ω RG-59/U cable with $a = 0.322$ mm (AWG 22-gauge wire), we find $b = 2$ mm, $C' = 68.5$ pF/m, $P_T = 49.5$ kW, and $E_{\max} = 0.46$ MV/m. □

Example 10.4.2: Most cables have a nominal impedance of either 50 or 75 Ω. The precise value depends on the manufacturer and the cable. For example, a 50-Ω cable might actually have an impedance of 52 Ω and a 75-Ω cable might actually be a 73-Ω cable.

The table below lists some commonly used cables with their AWG-gauge number of the inner conductor, the inner conductor radius a in mm, and their nominal impedance. Their dielectric filling is polyethylene with $\epsilon_r = 2.25$ or $n = \sqrt{\epsilon_r} = 1.5$.

type	AWG	a	Z
RG-6/U	18	0.512	75
RG-8/U	11	1.150	50
RG-11/U	14	0.815	75
RG-58/U	20	0.406	50
RG-59/U	22	0.322	75
RG-174/U	26	0.203	50
RG-213/U	13	0.915	50

The most commonly used cables are 50-Ω ones, such as the RG-58/U. Home cable-TV uses 75-Ω cables, such as the RG-59/U or RG-6/U.

The thin ethernet computer network, known as 10base-2, uses RG-58/U or RG-58A/U, which is similar to the RG-58/U but has a stranded inner copper core. Thick ethernet (10base-5) uses the thicker RG-8/U cable.

Because a dipole antenna has an input impedance of about 73 Ω, the RG-11, RG-6, and RG-59 75-Ω cables can be used to feed the antenna. □

Next, we determine the attenuation coefficient due to conductor losses. The power loss per unit length is given by Eq. (10.1.14). The magnetic fields at the surfaces of conductors (a) and (b) are obtained from Eq. (10.4.5) by setting $\rho = a$ and $\rho = b$:

$$H_a = \frac{I}{2\pi a}\,,\quad H_b = \frac{I}{2\pi b}$$

Because these are independent of the azimuthal angle, the integrations around the peripheries $dl = ad\phi$ or $dl = bd\phi$ will contribute a factor of $(2\pi a)$ or $(2\pi b)$. Thus,

$$P'_{\rm loss} = \frac{1}{2}R_s\big[(2\pi a)|H_a|^2 + (2\pi b)|H_b|^2\big] = \frac{R_s|I|^2}{4\pi}\left(\frac{1}{a}+\frac{1}{b}\right) \tag{10.4.7}$$

It follows that:

$$\alpha_c = \frac{P'_{\rm loss}}{2P_T} = \frac{\dfrac{R_s|I|^2}{4\pi}\left(\dfrac{1}{a}+\dfrac{1}{b}\right)}{2\dfrac{1}{2}Z|I|^2}$$

Using Eq. (10.4.4), we finally obtain:

$$\boxed{\alpha_c = \frac{R_s}{2\eta}\,\frac{\left(\dfrac{1}{a}+\dfrac{1}{b}\right)}{\ln\left(\dfrac{b}{a}\right)}} \tag{10.4.8}$$

The ohmic losses in the dielectric are described by Eq. (10.1.17). The total attenuation constant will be the sum of the conductor and dielectric attenuations:

$$\boxed{\alpha = \alpha_c + \alpha_d = \frac{R_s}{2\eta}\,\frac{\left(\dfrac{1}{a}+\dfrac{1}{b}\right)}{\ln\left(\dfrac{b}{a}\right)} + \frac{\omega}{2c}\tan\delta} \quad \text{(attenuation)} \tag{10.4.9}$$

The attenuation in dB/m will be $\alpha_{\rm dB} = 8.686\,\alpha$. This expression tends to somewhat underestimate the actual losses, but it is generally a good approximation. The α_c term grows in frequency like $\sqrt{f}$ and the term α_d, like f.

The smaller the dimensions a, b, the larger the attenuation. The loss tangent $\tan\delta$ of a typical polyethylene or teflon dielectric is of the order of 0.0004–0.0009 up to about 3 GHz.

The ohmic losses and the resulting heating of the dielectric and conductors also limit the power rating of the line. For example, if the maximum supported voltage is 1900 volts as in Example 10.4.2, the RMS value of the current for an RG-58/U line would be $I_{\rm rms} = 1900/53.5 = 35.5$ amps, which would melt the conductors. Thus, the actual power rating is much smaller than that suggested by the maximum voltage rating. The typical power rating of an RG-58/U cable is 1 kW, 200 W, and 80 W at 10 MHz, 200 MHz, and 1 GHz.

Example 10.4.3: The table below lists the nominal attenuations in dB per 100 feet of the RG-8/U and RG-213/U cables. The data are from [1322].

f (MHz)	50	100	200	400	900	1000	3000	5000
α (dB/100ft)	1.3	1.9	2.7	4.1	7.5	8.0	16.0	27.0

Both are 50-ohm cables and their radii a are 1.15 mm and 0.915 mm for RG-8/U and RG-213/U. In order to compare these ratings with Eq. (10.4.9), we took a to be the average of these two values, that is, $a = 1.03$ mm. The required value of b to give a 50-ohm impedance is $b = 3.60$ mm.

Fig. 10.4.2 shows the attenuations calculated from Eq. (10.4.9) and the nominal ones from the table. We assumed copper conductors with $\sigma = 5.8\times10^{7}$ S/m and polyethylene dielectric with $n = 1.5$, so that $\eta = \eta_0/n = 376.73/1.5 = 251.15$ Ω and $c = c_0/n = 2\times10^{8}$ m/sec. The loss tangent was taken to be $\tan\delta = 0.0007$.

The conductor and dielectric attenuations α_c and α_d become equal around 2.3 GHz, and α_d dominates after that.

It is evident that the useful operation of the cable is restricted to frequencies up to 1 GHz. Beyond that, the attenuations are too excessive and the cable may be used only for short lengths. □

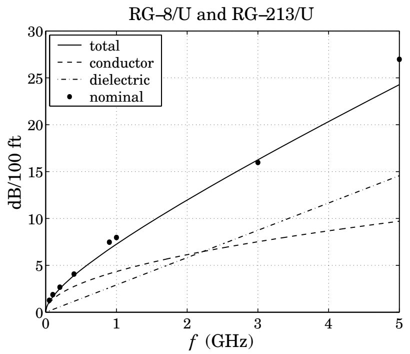


Fig. 10.4.2 Attenuation coefficient α versus frequency.

Optimum Coaxial Cables

Given a fixed outer-conductor radius b, one may ask three optimization questions: What is the optimum value of a, or equivalently, the ratio b/a that (a) minimizes the electric field E_a inside the guide, (b) maximizes the power transfer P_T, and (c) minimizes the conductor attenuation α_c.

The three quantities E_a, P_T, α_c can be thought of as functions of the ratio $x = b/a$ and take the following forms:

$$E_a = \frac{V}{b}\frac{x}{\ln x}\,,\quad P_T = \frac{1}{\eta}|E_a|^2\pi b^2\frac{\ln x}{x^2}\,,\quad \alpha_c = \frac{R_s}{2\eta b}\frac{x+1}{\ln x} \tag{10.4.10}$$

Setting the derivatives of the three functions of x to zero, we obtain the three conditions: (a) $\ln x = 1$, (b) $\ln x = 1/2$, and (c) $\ln x = 1 + 1/x$, with solutions (a) $b/a = e^1 = 2.7183$, (b) $b/a = e^{1/2} = 1.6487$ and (c) $b/a = 3.5911$.

Unfortunately, the three optimization problems have three different answers, and it is not possible to satisfy them simultaneously. The corresponding impedances Z for the three values of b/a are 60 Ω, 30 Ω, and 76.7 Ω for an air-filled line and 40 Ω, 20 Ω, and 51 Ω for a polyethylene-filled line.

The value of 50 Ω is considered to be a compromise between 30 and 76.7 Ω corresponding to maximum power and minimum attenuation. Actually, the minimum of α_c is very broad and any neighboring value to $b/a = 3.5911$ will result in an α_c very near its minimum.

Higher Modes

The TEM propagation mode is the dominant one and has no cutoff frequency. However, TE and TM modes with higher cutoff frequencies also exist in coaxial lines [838], with the lowest being a TE_{11} mode with cutoff wavelength and frequency:

$$\lambda_c = 1.873\frac{\pi}{2}(a+b)\,,\quad f_c = \frac{c}{\lambda_c} = \frac{c_0}{n\lambda_c} \tag{10.4.11}$$

This is usually approximated by $\lambda_c = \pi(a+b)$. Thus, the operation of the TEM mode is restricted to frequencies that are less than f_c.

Example 10.4.4: For the RG-58/U line of Example 10.4.2, we have $a = 0.406$ mm and $b = 1.548$ mm, resulting in $\lambda_c = 1.873\pi(a+b)/2 = 5.749$ mm, which gives for the cutoff frequency $f_c = 20/0.5749 = 34.79$ GHz, where we used $c = c_0/n = 20$ GHz cm.

For the RG-8/U and RG-213/U cables, we may use $a = 1.03$ mm and $b = 3.60$ as in Example 10.4.3, resulting in $\lambda_c = 13.622$ mm, and cutoff frequency of $f_c = 14.68$ GHz.

The above cutoff frequencies are far above the useful operating range over which the attenuation of the line is acceptable. □

10.5 Two-Wire Lines

The two-wire transmission line consists of two parallel cylindrical conductors of radius a separated by distance d from each other, as shown in Fig. 10.5.1.

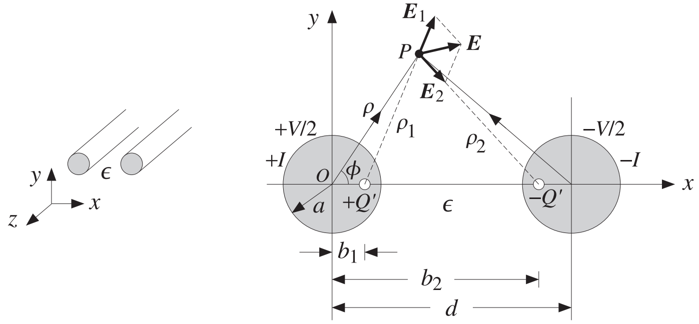


Fig. 10.5.1 Two-wire transmission line.

We assume that the conductors are held at potentials $\pm V/2$ with charge per unit length $\pm Q'$. The electrostatic problem can be solved by the standard technique of replacing the finite-radius conductors by two thin line-charges $\pm Q'$.

The locations b_1 and b_2 of the line-charges are determined by the requirement that the cylindrical surfaces of the original conductors be equipotential surfaces, the idea being that if these equipotential surfaces were to be replaced by the conductors, the field patterns will not be disturbed.

The electrostatic problem of the two lines is solved by invoking superposition and adding the potentials due to the two lines, so that the potential at the field point P will be:

$$\varphi(\rho,\phi) = -\frac{Q'}{2\pi\epsilon}\ln\rho_1 - \frac{-Q'}{2\pi\epsilon}\ln\rho_2 = \frac{Q'}{2\pi\epsilon}\ln\left(\frac{\rho_2}{\rho_1}\right) \tag{10.5.1}$$

where the ρ_1, ρ_2 are the distances from the line charges to P. From the triangles $OP(+Q')$ and $OP(-Q')$, we may express these distances in terms of the polar coordinates ρ, ϕ of the point P:

$$\rho_1 = \sqrt{\rho^2 - 2\rho b_1\cos\phi + b_1^2}\,,\qquad \rho_2 = \sqrt{\rho^2 - 2\rho b_2\cos\phi + b_2^2} \tag{10.5.2}$$

Therefore, the potential function becomes:

$$\varphi(\rho,\phi) = \frac{Q'}{2\pi\epsilon}\ln\left(\frac{\rho_2}{\rho_1}\right) = \frac{Q'}{2\pi\epsilon}\ln\left(\sqrt{\frac{\rho^2 - 2\rho b_2\cos\phi + b_2^2}{\rho^2 - 2\rho b_1\cos\phi + b_1^2}}\right) \tag{10.5.3}$$

In order that the surface of the left conductor at $\rho = a$ be an equipotential surface, that is, $\varphi(a,\phi) = V/2$, the ratio ρ_2/ρ_1 must be a constant independent of ϕ. Thus, we require that for some constant k and all angles ϕ:

$$\left.\frac{\rho_2}{\rho_1}\right|_{\rho=a} = \sqrt{\frac{a^2 - 2ab_2\cos\phi + b_2^2}{a^2 - 2ab_1\cos\phi + b_1^2}} = k$$

which can be rewritten as:

$$a^2 - 2ab_2\cos\phi + b_2^2 = k^2(a^2 - 2ab_1\cos\phi + b_1^2)$$

This will be satisfied for all ϕ provided we have:

$$a^2 + b_2^2 = k^2(a^2 + b_1^2)\,,\qquad b_2 = k^2 b_1$$

These may be solved for b_1, b_2 in terms of k:

$$b_2 = ka\,,\quad b_1 = \frac{a}{k} \tag{10.5.4}$$

The quantity k can be expressed in terms of a, d by noting that because of symmetry, the charge $-Q'$ is located also at distance b_1 from the center of the right conductor. Therefore, $b_1 + b_2 = d$. This gives the condition:

$$b_1 + b_2 = d \quad\Rightarrow\quad a(k + k^{-1}) = d \quad\Rightarrow\quad k + k^{-1} = \frac{d}{a}$$

with solution for k:

$$k = \frac{d}{2a} + \sqrt{\left(\frac{d}{2a}\right)^2 - 1} \tag{10.5.5}$$

An alternative expression is obtained by setting $k = e^{\chi}$. Then, we have the condition:

$$b_1 + b_2 = d \quad\Rightarrow\quad a(e^{\chi} + e^{-\chi}) = 2a\cosh\chi = d \quad\Rightarrow\quad \chi = \operatorname{acosh}\left(\frac{d}{2a}\right) \tag{10.5.6}$$

Because $\chi = \ln k$, we obtain for the potential value of the left conductor:

$$\varphi(a,\phi) = \frac{Q'}{2\pi\epsilon}\ln k = \frac{Q'}{2\pi\epsilon}\chi = \frac{1}{2}V$$

This gives for the capacitance per unit length:

$$\boxed{C' = \frac{Q'}{V} = \frac{\pi\epsilon}{\chi} = \frac{\pi\epsilon}{\operatorname{acosh}\left(\dfrac{d}{2a}\right)}} \tag{10.5.7}$$

The corresponding line impedance and inductance are obtained from $C' = \epsilon\eta/Z$ and $L' = \mu Z/\eta$. We find:

$$\boxed{Z = \frac{\eta}{\pi}\chi = \frac{\eta}{\pi}\operatorname{acosh}\left(\frac{d}{2a}\right)}\qquad \boxed{L' = \frac{\mu}{\pi}\chi = \frac{\mu}{\pi}\operatorname{acosh}\left(\frac{d}{2a}\right)} \tag{10.5.8}$$

In the common case when $d \gg a$, we have approximately $k \simeq d/a$, and therefore, $\chi = \ln k = \ln(d/a)$. Then, Z can be written approximately as:

$$Z = \frac{\eta}{\pi}\ln(d/a) \tag{10.5.9}$$

To complete the electrostatic problem and determine the electric and magnetic fields of the TEM mode, we replace $b_2 = ak$ and $b_1 = a/k$ in Eq. (10.5.3) and write it as:

$$\varphi(\rho,\phi) = \frac{Q'}{2\pi\epsilon}\ln\left(k\sqrt{\frac{\rho^2 - 2ak\rho\cos\phi + a^2k^2}{\rho^2k^2 - 2ak\rho\cos\phi + a^2}}\right) \tag{10.5.10}$$

The electric and magnetic field components are obtained from:

$$E_\rho = \eta H_\phi = -\frac{\partial\varphi}{\partial\rho}\,,\quad E_\phi = -\eta H_\rho = -\frac{\partial\varphi}{\rho\partial\phi} \tag{10.5.11}$$

Performing the differentiations, we find:

$$\begin{aligned}
E_\rho &= -\frac{Q'}{2\pi\epsilon}\left[\frac{\rho - ak\cos\phi}{\rho^2 - 2ak\rho\cos\phi + a^2k^2} - \frac{\rho k^2 - ak\cos\phi}{\rho^2k^2 - 2ak\rho\cos\phi + a^2}\right]\\
E_\phi &= -\frac{Q'}{2\pi\epsilon}\left[\frac{ak\sin\phi}{\rho^2 - 2ak\cos\phi + a^2k^2} - \frac{ak\rho\sin\phi}{\rho^2k^2 - 2ak\rho\cos\phi + a^2}\right]
\end{aligned} \tag{10.5.12}$$

The resistance per unit length and corresponding attenuation constant due to conductor losses are calculated in Problem 10.3:

$$R' = \frac{R_s}{\pi a}\,\frac{d}{\sqrt{d^2 - 4a^2}}\,,\qquad \alpha_c = \frac{R'}{2Z} = \frac{R_s}{2\eta a}\,\frac{d}{\operatorname{acosh}(d/2a)\sqrt{d^2 - 4a^2}} \tag{10.5.13}$$

10.6 Distributed Circuit Model of a Transmission Line

We saw that a transmission line has associated with it the parameters L', C' describing its lossless operation, and in addition, the parameters R', G' which describe the losses. It is possible then to define a series impedance Z' and a shunt admittance Y' per unit length by combining R' with L' and G' with C':

$$\boxed{\begin{aligned} Z' &= R' + j\omega L' \\ Y' &= G' + j\omega C' \end{aligned}} \tag{10.6.1}$$

This leads to a so-called *distributed-parameter* circuit, which means that every infinitesimal segment Δz of the line can be replaced by a series impedance $Z'\Delta z$ and a shunt admittance $Y'\Delta z$, as shown in Fig. 10.6.1. The voltage and current at location z will be $V(z), I(z)$ and at location $z+\Delta z$, $V(z+\Delta z), I(z+\Delta z)$.

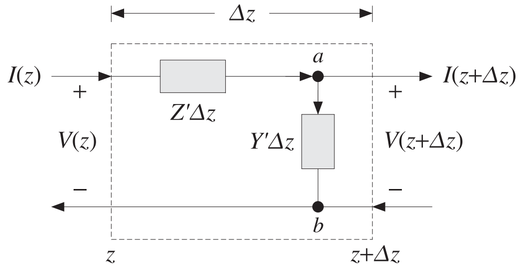


Fig. 10.6.1 Distributed parameter model of a transmission line.

The voltage across the branch a–b is $V_{ab} = V(z+\Delta z)$ and the current through it, $I_{ab} = (Y'\Delta z)V_{ab} = Y'\Delta z\, V(z+\Delta z)$. Applying Kirchhoff's voltage and current laws, we obtain:

$$\begin{aligned} V(z) &= (Z'\Delta z)\, I(z) + V_{ab} = Z'\Delta z\, I(z) + V(z+\Delta z) \\ I(z) &= I_{ab} + I(z+\Delta z) = Y'\Delta z\, V(z+\Delta z) + I(z+\Delta z) \end{aligned} \tag{10.6.2}$$

Using a Taylor series expansion, we may expand $I(z+\Delta z)$ and $V(z+\Delta z)$ to first order in Δz:

$$\begin{aligned} I(z+\Delta z) &= I(z) + I'(z)\Delta z \\ V(z+\Delta z) &= V(z) + V'(z)\Delta z \quad \text{and} \quad Y'\Delta z\, V(z+\Delta z) = Y'\Delta z\, V(z) \end{aligned}$$

Inserting these expressions in Eq. (10.6.2) and matching the zeroth- and first-order terms in the two sides, we obtain the equivalent differential equations:

$$\boxed{\begin{aligned} V'(z) &= -Z'I(z) = -(R' + j\omega L')I(z) \\ I'(z) &= -Y'V(z) = -(G' + j\omega C')V(z) \end{aligned}} \tag{10.6.3}$$

It is easily verified that the most general solution of this coupled system is expressible as a sum of a forward and a backward moving wave:

$$\begin{aligned} V(z) &= V_+ e^{-j\beta_c z} + V_- e^{j\beta_c z} \\ I(z) &= \frac{1}{Z_c}\big(V_+ e^{-j\beta_c z} - V_- e^{j\beta_c z}\big) \end{aligned} \tag{10.6.4}$$

where β_c, Z_c are the complex wavenumber and complex impedance:

$$\begin{aligned} \beta_c &= -j\sqrt{(R' + j\omega L')(G' + j\omega C')} = \omega\sqrt{L'C'}\sqrt{\left(1 - j\frac{R'}{\omega L'}\right)\left(1 - j\frac{G'}{\omega C'}\right)} \\ Z_c &= \sqrt{\frac{Z'}{Y'}} = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \end{aligned} \tag{10.6.5}$$

The time-domain impulse response of such a line was given in Sec. 3.3. The real and imaginary parts of $\beta_c = \beta - j\alpha$ define the propagation and attenuation constants. In the case of a lossless line, $R' = G' = 0$, we obtain using Eq. (10.1.6):

$$\beta_c = \omega\sqrt{L'C'} = \omega\sqrt{\mu\epsilon} = \frac{\omega}{c} = \beta\,, \quad Z_c = \sqrt{\frac{L'}{C'}} = Z \tag{10.6.6}$$

In practice, we always assume a lossless line and then take into account the losses by assuming that R' and G' are small quantities, which can be evaluated by the appropriate expressions that can be derived for each type of line, as we did for the parallel-plate, coaxial, and two-wire lines. The lossless solution (10.6.4) takes the form:

$$\boxed{\begin{aligned} V(z) &= V_+ e^{-j\beta z} + V_- e^{j\beta z} = V_+(z) + V_-(z) \\ I(z) &= \frac{1}{Z}\big(V_+ e^{-j\beta z} - V_- e^{j\beta z}\big) = \frac{1}{Z}\big(V_+(z) - V_-(z)\big) \end{aligned}} \tag{10.6.7}$$

This solution is identical to that of uniform plane waves of Chap. 5, provided we make the identifications:

$$\boxed{\begin{aligned} V(z) &\longleftrightarrow E(z) \\ I(z) &\longleftrightarrow H(z) \\ Z &\longleftrightarrow \eta \end{aligned}} \quad \text{and} \quad \boxed{\begin{aligned} V_+(z) &\longleftrightarrow E_+(z) \\ V_-(z) &\longleftrightarrow E_-(z) \end{aligned}}$$

10.7 Wave Impedance and Reflection Response

All the concepts of Chap. 5 translate verbatim to the transmission line case. For example, we may define the *wave impedance* and *reflection response* at location z:

$$Z(z) = \frac{V(z)}{I(z)} = Z_0\frac{V_+(z) + V_-(z)}{V_+(z) - V_-(z)}\,, \quad \Gamma(z) = \frac{V_-(z)}{V_+(z)} \tag{10.7.1}$$

To avoid ambiguity in notation, we will denote the characteristic impedance of the line by Z_0. It follows from Eq. (10.7.1) that $Z(z)$ and $\Gamma(z)$ are related by:

$$\boxed{Z(z) = Z_0\frac{1 + \Gamma(z)}{1 - \Gamma(z)}}\,, \quad \boxed{\Gamma(z) = \frac{Z(z) - Z_0}{Z(z) + Z_0}} \tag{10.7.2}$$

For a forward-moving wave, the conditions $\Gamma(z) = 0$ and $Z(z) = Z_0$ are equivalent. The propagation equations of $Z(z)$ and $\Gamma(z)$ between two points z_1, z_2 along the line separated by distance $l = z_2 - z_1$ are given by:

$$\boxed{Z_1 = Z_0\frac{Z_2 + jZ_0\tan\beta l}{Z_0 + jZ_2\tan\beta l}} \quad\Leftrightarrow\quad \boxed{\Gamma_1 = \Gamma_2 e^{-2j\beta l}} \tag{10.7.3}$$

where we have the relationships between Z_1, Z_2 and Γ_1, Γ_2:

$$Z_1 = Z_0\frac{1+\Gamma_1}{1-\Gamma_1}\,, \quad Z_2 = Z_0\frac{1+\Gamma_2}{1-\Gamma_2} \tag{10.7.4}$$

We may also express Z_1 in terms of Γ_2:

$$\boxed{Z_1 = Z_0\frac{1+\Gamma_1}{1-\Gamma_1} = Z_0\frac{1+\Gamma_2 e^{-2j\beta l}}{1-\Gamma_2 e^{-2j\beta l}}} \tag{10.7.5}$$

The relationship between the voltage and current waves at points z_1 and z_2 is obtained by the propagation matrix:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} \cos\beta l & jZ_0\sin\beta l \\ jZ_0^{-1}\sin\beta l & \cos\beta l \end{bmatrix}\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \quad \text{(propagation matrix)} \tag{10.7.6}$$

Similarly, we may relate the forward/backward voltages at the points z_1 and z_2:

$$\begin{bmatrix} V_{1+} \\ V_{1-} \end{bmatrix} = \begin{bmatrix} e^{j\beta l} & 0 \\ 0 & e^{-j\beta l} \end{bmatrix}\begin{bmatrix} V_{2+} \\ V_{2-} \end{bmatrix} \quad \text{(propagation matrix)} \tag{10.7.7}$$

It follows from Eq. (10.6.7) that $V_{1\pm}, V_{2\pm}$ are related to V_1, I_1 and V_2, I_2 by:

$$V_{1\pm} = \frac{1}{2}(V_1 \pm Z_0 I_1)\,, \quad V_{2\pm} = \frac{1}{2}(V_2 \pm Z_0 I_2) \tag{10.7.8}$$

Fig. 10.7.1 depicts these various quantities. We note that the behavior of the line remains unchanged if the line is cut at the point z_2 and the entire right portion of the line is replaced by an impedance equal to Z_2, as shown in the figure.

This is so because in both cases, all the points z_1 to the left of z_2 see the same voltage-current relationship at z_2, that is, $V_2 = Z_2 I_2$.

Sometimes, as in the case of designing stub tuners for matching a line to a load, it is more convenient to work with the *wave admittances*. Defining $Y_0 = 1/Z_0$, $Y_1 = 1/Z_1$, and $Y_2 = 1/Z_2$, it is easily verified that the admittances satisfy exactly the same propagation relationship as the impedances:

$$\boxed{Y_1 = Y_0\frac{Y_2 + jY_0\tan\beta l}{Y_0 + jY_2\tan\beta l}} \tag{10.7.9}$$

As in the case of dielectric slabs, the half- and quarter-wavelength separations are of special interest. For a half-wave distance, we have $\beta l = 2\pi/2 = \pi$, which translates to $l = \lambda/2$, where $\lambda = 2\pi/\beta$ is the wavelength along the line. For a quarter-wave, we have $\beta l = 2\pi/4 = \pi/2$ or $l = \lambda/4$. Setting $\beta l = \pi$ or $\pi/2$ in Eq. (10.7.3), we obtain:

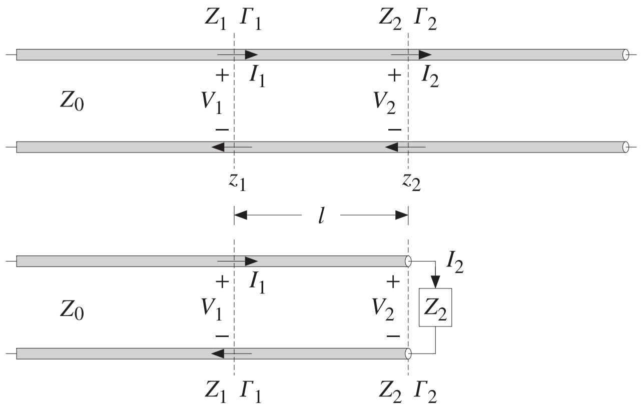


Fig. 10.7.1 Length segment on infinite line and equivalent terminated line.

$$\boxed{\begin{array}{lll} l = \dfrac{\lambda}{2} & \Rightarrow & Z_1 = Z_2, \quad \Gamma_1 = \Gamma_2 \\ l = \dfrac{\lambda}{4} & \Rightarrow & Z_1 = \dfrac{Z_0^2}{Z_2}, \quad \Gamma_1 = -\Gamma_2 \end{array}} \tag{10.7.10}$$

The MATLAB functions `z2g.m` and `g2z.m` compute Γ from Z and conversely, by implementing Eq. (10.7.2). The functions `gprop.m`, `zprop.m` and `vprop.m` implement the propagation equations (10.7.3) and (10.7.6). The usage of these functions is:

```
G  = z2g(Z,Z0);                     % Z to Γ
Z  = g2z(G,Z0);                     % Γ to Z
G1 = gprop(G2,bl);                  % propagates Γ2 to Γ1
Z1 = zprop(Z2,Z0,bl);               % propagates Z2 to Z1
[V1,I1] = vprop(V2,I2,Z0,bl);       % propagates V2,I2 to V1,I1
```

The parameter `bl` is βl. The propagation equations and these MATLAB functions also work for lossy lines. In this case, β must be replaced by the complex wavenumber $\beta_c = \beta - j\alpha$. The propagation phase factors become now:

$$e^{\pm j\beta l} \longrightarrow e^{\pm j\beta_c l} = e^{\pm\alpha l}e^{\pm j\beta l} \tag{10.7.11}$$

10.8 Two-Port Equivalent Circuit

Any length-l segment of a transmission line may be represented as a two-port equivalent circuit. Rearranging the terms in Eq. (10.7.6), we may write it in *impedance-matrix* form:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}\begin{bmatrix} I_1 \\ -I_2 \end{bmatrix} \quad \text{(impedance matrix)} \tag{10.8.1}$$

where the impedance elements are:

$$
\begin{aligned}
Z_{11} &= Z_{22} = -jZ_0 \cot\beta l \\
Z_{12} &= Z_{21} = -jZ_0 \frac{1}{\sin\beta l}
\end{aligned}
\tag{10.8.2}
$$

The negative sign, $-I_2$, conforms to the usual convention of having the currents coming *into* the two-port from either side. This impedance matrix can also be realized in a T-section configuration as shown in Fig. 10.8.1.

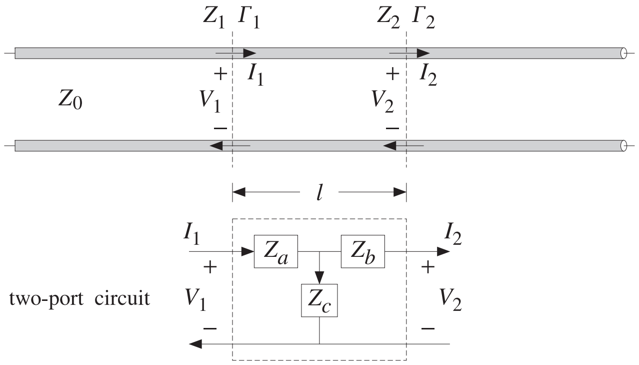


Fig. 10.8.1 Length-l segment of a transmission line and its equivalent T-section.

Using Eq. (10.8.1) and some trigonometry, the impedances Z_a, Z_b, Z_c of the T-section are found to be:

$$
\begin{aligned}
Z_a &= Z_{11} - Z_{12} = jZ_0 \tan(\beta l/2) \\
Z_b &= Z_{22} - Z_{12} = jZ_0 \tan(\beta l/2) \\
Z_c &= Z_{12} = -jZ_0 \frac{1}{\sin\beta l}
\end{aligned}
\tag{10.8.3}
$$

The MATLAB function `tsection.m` implements Eq. (10.8.3). Its usage is:

```
[Za,Zc] = tsection(Z0,bl);
```

10.9 Terminated Transmission Lines

We can use the results of the previous section to analyze the behavior of a transmission line connected between a generator and a load. For example in a transmitting antenna system, the transmitter is the generator and the antenna, the load. In a receiving system, the antenna is the generator and the receiver, the load.

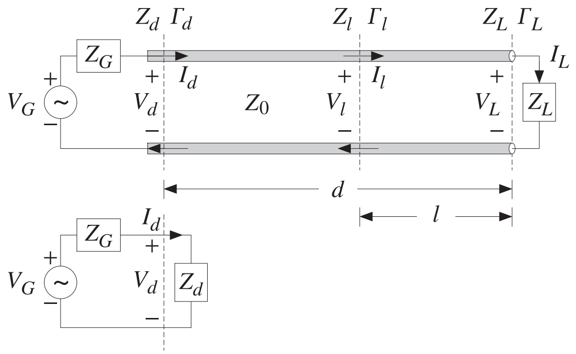


Fig. 10.9.1 Terminated line and equivalent circuit.

Fig. 10.9.1 shows a generator of voltage V_G and internal impedance Z_G connected to the load impedance Z_L through a length d of a transmission line of characteristic impedance Z_0. We wish to determine the voltage and current at the load in terms of the generator voltage.

We assume that the line is lossless and hence Z_0 is real. The generator impedance is also assumed to be real but it does not have to be. The load impedance will have in general both a resistive and a reactive part, $Z_L = R_L + jX_L$.

At the load location, the voltage, current, and impedance are V_L, I_L, Z_L and play the same role as the quantities V_2, I_2, Z_2 of the previous section. They are related by $V_L = Z_L I_L$. The reflection coefficient at the load will be:

$$
\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \quad \Leftrightarrow \quad Z_L = Z_0 \frac{1+\Gamma_L}{1-\Gamma_L}
\tag{10.9.1}
$$

The quantities Z_L, Γ_L can be propagated now by a distance d to the generator at the input to the line. The corresponding voltage, current, and impedance V_d, I_d, Z_d play the role of V_1, I_1, Z_1 of the previous section, and are related by $V_d = Z_d I_d$. We have the propagation relationships:

$$
Z_d = Z_0 \frac{Z_L + jZ_0 \tan\beta d}{Z_0 + jZ_L \tan\beta d} \quad \Leftrightarrow \quad \Gamma_d = \Gamma_L e^{-2j\beta d}
\tag{10.9.2}
$$

where

$$
\Gamma_d = \frac{Z_d - Z_0}{Z_d + Z_0} \quad \Leftrightarrow \quad Z_d = Z_0 \frac{1+\Gamma_d}{1-\Gamma_d} = Z_0 \frac{1+\Gamma_L e^{-2j\beta d}}{1-\Gamma_L e^{-2j\beta d}}
\tag{10.9.3}
$$

At the line input, the entire length-d line segment and load can be replaced by the impedance Z_d, as shown in Fig. 10.9.1. We have now a simple voltage divider circuit. Thus,

$$
\boxed{V_d = V_G - I_d Z_G = \frac{V_G Z_d}{Z_G + Z_d}\,, \quad I_d = \frac{V_G}{Z_G + Z_d}}
\tag{10.9.4}
$$

Once we have V_d, I_d in terms of V_G, we can invert the propagation matrix (10.7.6) to obtain the voltage and current at the load:

$$\begin{bmatrix} V_L \\ I_L \end{bmatrix} = \begin{bmatrix} \cos\beta d & -jZ_0\sin\beta d \\ -jZ_0^{-1}\sin\beta d & \cos\beta d \end{bmatrix} \begin{bmatrix} V_d \\ I_d \end{bmatrix} \tag{10.9.5}$$

It is more convenient to express V_d, I_d in terms of the reflection coefficients Γ_d and Γ_G, the latter being defined by:

$$\Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0} \quad \Leftrightarrow \quad Z_G = Z_0\,\frac{1+\Gamma_G}{1-\Gamma_G} \tag{10.9.6}$$

It is easy to verify using Eqs. (10.9.3) and (10.9.6) that:

$$Z_G + Z_d = 2Z_0\,\frac{1-\Gamma_G\Gamma_d}{(1-\Gamma_G)(1-\Gamma_d)}\,, \quad Z_G + Z_0 = 2Z_0\,\frac{1}{1-\Gamma_G}$$

From these, it follows that:

$$\boxed{V_d = \frac{V_G Z_0}{Z_G + Z_0}\,\frac{1+\Gamma_d}{1-\Gamma_G\Gamma_d}\,, \quad I_d = \frac{V_G}{Z_G + Z_0}\,\frac{1-\Gamma_d}{1-\Gamma_G\Gamma_d}} \tag{10.9.7}$$

where Γ_d may be replaced by $\Gamma_d = \Gamma_L e^{-2j\beta d}$. If the line and load are matched so that $Z_L = Z_0$, then $\Gamma_L = 0$ and $\Gamma_d = 0$ and $Z_d = Z_0$ for any distance d. Eq. (10.9.7) then reduces to:

$$\boxed{V_d = \frac{V_G Z_0}{Z_G + Z_0}\,, \quad I_d = \frac{V_G}{Z_G + Z_0}} \qquad \text{(matched load)} \tag{10.9.8}$$

In this case, there is only a forward-moving wave along the line. The voltage and current at the load will correspond to the propagation of these quantities to location $l = 0$, which introduces a propagation phase factor $e^{-j\beta d}$:

$$\boxed{V_0 = \frac{V_G Z_0}{Z_G + Z_0}\,e^{-j\beta d}\,, \quad I_0 = \frac{V_G}{Z_G + Z_0}\,e^{-j\beta d}} \qquad \text{(matched load)} \tag{10.9.9}$$

where V_0, I_0 denote V_L, I_L when $Z_L = Z_0$. It is convenient also to express V_L directly in terms of V_d and the reflection coefficients Γ_d and Γ_L. We note that:

$$V_L = V_{L+}(1+\Gamma_L)\,, \quad V_{L+} = V_{d+}e^{-j\beta d}\,, \quad V_{d+} = \frac{V_d}{1+\Gamma_d}$$

It follows that the voltage V_L and current $I_L = V_L/Z_L$ are:

$$\boxed{V_L = V_d e^{-j\beta d}\,\frac{1+\Gamma_L}{1+\Gamma_d}\,, \quad I_L = I_d e^{-j\beta d}\,\frac{1-\Gamma_L}{1-\Gamma_d}} \tag{10.9.10}$$

Expressing V_L and also $I_L = V_L/Z_L$ directly in terms of V_G, we have:

$$\boxed{V_L = \frac{V_G Z_0}{Z_G + Z_0}\,\frac{1+\Gamma_L}{1-\Gamma_G\Gamma_d}\,e^{-j\beta d}\,, \quad I_L = \frac{V_G}{Z_G + Z_0}\,\frac{1-\Gamma_L}{1-\Gamma_G\Gamma_d}\,e^{-j\beta d}} \tag{10.9.11}$$

It should be emphasized that d refers to the fixed distance between the generator and the load. For any other distance, say l, from the load (or, distance $z = d - l$ from the generator,) the voltage and current can be expressed in terms of the load voltage and current as follows:

$$\boxed{V_l = V_L e^{j\beta l}\,\frac{1+\Gamma_l}{1+\Gamma_L}\,, \quad I_l = I_L e^{j\beta l}\,\frac{1-\Gamma_l}{1-\Gamma_L}}\,, \quad \Gamma_l = \Gamma_L e^{-2j\beta l} \tag{10.9.12}$$

10.10 Power Transfer from Generator to Load

The *total power* delivered by the generator is dissipated partly in its internal resistance and partly in the load. The power delivered to the load is equal (for a lossless line) to the net power traveling to the right at any point along the line. Thus, we have:

$$P_{\text{tot}} = P_d + P_G = P_L + P_G \tag{10.10.1}$$

This follows from $V_G = V_d + I_d Z_G$, which implies

$$V_G I_d^* = V_d I_d^* + Z_G|I_d|^2 \tag{10.10.2}$$

Eq. (10.10.1) is a consequence of (10.10.2) and the definitions:

$$\begin{aligned} P_{\text{tot}} &= \frac{1}{2}\,\text{Re}(V_G^* I_d) = \frac{1}{2}\,\text{Re}\big[(V_d + Z_G I_d)^* I_d\big] \\ P_G &= \frac{1}{2}\,\text{Re}(Z_G I_d I_d^*) = \frac{1}{2}\,\text{Re}(Z_G)\,|I_d|^2 \\ P_d &= \frac{1}{2}\,\text{Re}(V_d^* I_d) = \frac{1}{2}\,\text{Re}(V_L^* I_L) = P_L \end{aligned} \tag{10.10.3}$$

The last equality follows from Eq. (10.9.5) or from $V_{d\pm} = V_{L\pm}e^{\pm j\beta d}$:

$$\frac{1}{2}\,\text{Re}(V_d^* I_d) = \frac{1}{2Z_0}\big(|V_{d+}|^2 - |V_{d-}|^2\big) = \frac{1}{2Z_0}\big(|V_{L+}|^2 - |V_{L-}|^2\big) = \frac{1}{2}\,\text{Re}(V_L^* I_L)$$

In the special case when the generator and the load are matched to the line, so that $Z_G = Z_L = Z_0$, then we find the standard result that half of the generated power is delivered to the load and half is lost in the internal impedance. Using Eq. (10.9.8) with $Z_G = Z_0$, we obtain $V_d = I_d Z_G = V_G/2$, which gives:

$$P_{\text{tot}} = \frac{|V_G|^2}{4Z_0}\,, \quad P_G = \frac{|V_G|^2}{8Z_0} = \frac{1}{2}P_{\text{tot}}\,, \quad P_d = P_L = \frac{|V_G|^2}{8Z_0} = \frac{1}{2}P_{\text{tot}} \tag{10.10.4}$$

Example 10.10.1: A load $Z_L = 50 + j10\ \Omega$ is connected to a generator $V_G = 10\angle 0^{\text{o}}$ volts with a 100-ft (30.48 m) cable of a 50-ohm transmission line. The generator's internal impedance is 20 ohm, the operating frequency is 10 MHz, and the velocity factor of the line, 2/3.

Determine the voltage across the load, the total power delivered by the generator, the power dissipated in the generator's internal impedance and in the load.

Solution: The propagation speed is $c = 2c_0/3 = 2\times 10^8$ m/sec. The line wavelength $\lambda = c/f = 20$ m and the propagation wavenumber $\beta = 2\pi/\lambda = 0.3142$ rads/m. The electrical length is $d/\lambda = 30.48/20 = 1.524$ and the phase length $\beta d = 9.5756$ radians.

Next, we calculate the reflection coefficients:

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} = 0.0995\angle 84.29^{\text{o}}\,,\quad \Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0} = -0.4286$$

and $\Gamma_d = \Gamma_L e^{-2j\beta d} = 0.0995\angle 67.01^{\text{o}}$. It follows that:

$$Z_d = Z_0\frac{1+\Gamma_d}{1-\Gamma_d} = 53.11 + j9.83\,,\quad V_d = \frac{V_G Z_d}{Z_G + Z_d} = 7.31 + j0.36 = 7.32\angle 2.83^{\text{o}}$$

The voltage across the load will be:

$$V_L = V_d e^{-j\beta d}\,\frac{1+\Gamma_L}{1+\Gamma_d} = -7.09 + j0.65 = 7.12\angle 174.75^{\text{o}}\ \text{V}$$

The current through the generator is:

$$I_d = \frac{V_d}{Z_d} = 0.13 - j0.02 = 0.14\angle -7.66^{\text{o}}\ \text{A}$$

It follows that the generated and dissipated powers will be:

$$\begin{aligned} P_{\text{tot}} &= \frac{1}{2}\,\text{Re}(V_G^* I_d) = 0.6718\ \text{W} \\ P_G &= \frac{1}{2}\,\text{Re}(Z_G)\,|I_d|^2 = 0.1838\ \text{W} \\ P_L = P_d &= \frac{1}{2}\,\text{Re}(V_d^* I_d) = 0.4880\ \text{W} \end{aligned}$$

We note that $P_{\text{tot}} = P_G + P_L$. □

If the line is lossy, with a complex wavenumber $\beta_c = \beta - j\alpha$, the power P_L at the output of the line is less than the power P_d at the input of the line. Writing $V_{d\pm} = V_{L\pm}e^{\pm\alpha d}e^{\pm j\beta d}$, we find:

$$\begin{aligned} P_d &= \frac{1}{2Z_0}\big(|V_{d+}|^2 - |V_{d-}|^2\big) = \frac{1}{2Z_0}\big(|V_{L+}|^2 e^{2\alpha d} - |V_{L-}|^2 e^{-2\alpha d}\big) \\ P_L &= \frac{1}{2Z_0}\big(|V_{L+}|^2 - |V_{L-}|^2\big) \end{aligned}$$

We note that $P_d > P_L$ for all Γ_L. In terms of the incident forward power at the load, $P_{\text{inc}} = |V_{L+}|^2/2Z_0$, we have:

$$\begin{aligned} P_d &= P_{\text{inc}}\big(e^{2\alpha d} - |\Gamma_L|^2 e^{-2\alpha d}\big) = P_{\text{inc}}e^{2\alpha d}\big(1 - |\Gamma_d|^2\big) \\ P_L &= P_{\text{inc}}\big(1 - |\Gamma_L|^2\big) \end{aligned} \tag{10.10.5}$$

where $|\Gamma_d| = |\Gamma_L|e^{-2\alpha d}$. The total attenuation or loss of the line is P_d/P_L (the inverse P_L/P_d is the total gain, which is less than one.) In decibels, the loss is:

$$\boxed{L = 10\log_{10}\left(\frac{P_d}{P_L}\right) = 10\log_{10}\left(\frac{e^{2\alpha d} - |\Gamma_L|^2 e^{-2\alpha d}}{1 - |\Gamma_L|^2}\right)} \qquad \text{(total loss)} \tag{10.10.6}$$

If the load is matched to the line, $Z_L = Z_0$, so that $\Gamma_L = 0$, the loss is referred to as the *matched-line* loss and is due only to the transmission losses along the line:

$$\boxed{L_M = 10\log_{10}\big(e^{2\alpha d}\big) = 8.686\alpha d} \qquad \text{(matched-line loss)} \tag{10.10.7}$$

Denoting the matched-line loss in absolute units by $a = 10^{L_M/10} = e^{2\alpha d}$, we may write Eq. (10.10.6) in the equivalent form:

$$\boxed{L = 10\log_{10}\left(\frac{a^2 - |\Gamma_L|^2}{a\big(1 - |\Gamma_L|^2\big)}\right)} \qquad \text{(total loss)} \tag{10.10.8}$$

The additional loss due to the mismatched load is the difference:

$$L - L_M = 10\log_{10}\left(\frac{1 - |\Gamma_L|^2 e^{-4\alpha d}}{1 - |\Gamma_L|^2}\right) = 10\log_{10}\left(\frac{1 - |\Gamma_d|^2}{1 - |\Gamma_L|^2}\right) \tag{10.10.9}$$

Example 10.10.2: A 150 ft long RG-58 coax is connected to a load $Z_L = 25 + 50j$ ohm. At the operating frequency of 10 MHz, the cable is rated to have 1.2 dB/100 ft of matched-line loss. Determine the total loss of the line and the excess loss due to the mismatched load.

Solution: The matched-line loss of the 150 ft cable is $L_M = 150\times 1.2/100 = 1.8$ dB or in absolute units, $a = 10^{1.8/10} = 1.51$. The reflection coefficient has magnitude computed with the help of the MATLAB function `z2g`:

$$|\Gamma_L| = \text{abs}\big(\text{z2g}(25 + 50j, 50)\big) = 0.62$$

It follows that the total loss will be:

$$L = 10\log_{10}\left(\frac{a^2 - |\Gamma_L|^2}{a\big(1 - |\Gamma_L|^2\big)}\right) = 10\log_{10}\left(\frac{1.51^2 - 0.62^2}{1.51\big(1 - 0.62^2\big)}\right) = 3.1\ \text{dB}$$

The excess loss due to the mismatched load is $3.1 - 1.8 = 1.3$ dB. At the line input, we have $|\Gamma_d| = |\Gamma_L|e^{-2\alpha d} = |\Gamma_L|/a = 0.62/1.51 = 0.41$. Therefore, from the point of view of the input the line appears to be more matched. □

10.11 Open- and Short-Circuited Transmission Lines

Open- and short-circuited transmission lines are widely used to construct resonant circuits as well as matching stubs. They correspond to the special cases for the load impedance: $Z_L = \infty$ for an open-circuited line and $Z_L = 0$ for a short-circuited one. Fig. 10.11.1 shows these two cases.

Knowing the open-circuit voltage and the short-circuit current at the end terminals a, b, allows us also to replace the entire left segment of the line, including the generator, with a Thévenin-equivalent circuit. Connected to a load impedance Z_L, the equivalent circuit will produce the same load voltage and current V_L, I_L as the original line and generator.

Setting $Z_L = \infty$ and $Z_L = 0$ in Eq. (10.9.2), we obtain the following expressions for the wave impedance Z_l at distance l from the open- or short-circuited termination:

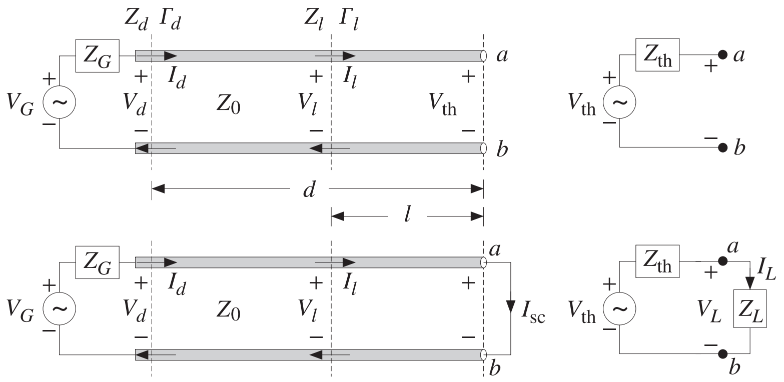


Fig. 10.11.1 Open- and short-circuited line and Thévenin-equivalent circuit.

$$\boxed{\begin{aligned} Z_l &= -jZ_0\cot\beta l \\ Z_l &= jZ_0\tan\beta l \end{aligned}} \quad \begin{aligned} &\text{(open-circuited)} \\ &\text{(short-circuited)} \end{aligned} \tag{10.11.1}$$

The corresponding admittances $Y_l = 1/Z_l$ will be:

$$\boxed{\begin{aligned} Y_l &= jY_0\tan\beta l \\ Y_l &= -jY_0\cot\beta l \end{aligned}} \quad \begin{aligned} &\text{(open-circuited)} \\ &\text{(short-circuited)} \end{aligned} \tag{10.11.2}$$

To determine the Thévenin-equivalent circuit that replaces everything to the left of the terminals a, b, we must find the open-circuit voltage $V_{\rm th}$, the short-circuit current $I_{\rm sc}$, and the Thévenin impedance $Z_{\rm th}$.

The impedance $Z_{\rm th}$ can be determined either by $Z_{\rm th} = V_{\rm th}/I_{\rm sc}$, or by disconnecting the generator and finding the equivalent impedance looking to the left of the terminals a, b. It is obtained by propagating the generator impedance Z_G by a distance d:

$$\boxed{Z_{\rm th} = Z_0\,\frac{Z_G + jZ_0\tan\beta d}{Z_0 + jZ_G\tan\beta d} = Z_0\,\frac{1+\Gamma_{\rm th}}{1-\Gamma_{\rm th}}}\,, \quad \Gamma_{\rm th} = \Gamma_G e^{-2j\beta d} \tag{10.11.3}$$

The open-circuit voltage can be determined from Eq. (10.9.11) by setting $Z_L = \infty$, which implies that $\Gamma_L = 1$, $\Gamma_d = e^{-2j\beta d}$, and $\Gamma_G\Gamma_d = \Gamma_G e^{-2j\beta d} = \Gamma_{\rm th}$. The short-circuit current is also obtained from (10.9.11) by setting $Z_L = 0$, which gives $\Gamma_L = -1$, $\Gamma_d = -e^{-2j\beta d}$, and $\Gamma_G\Gamma_d = -\Gamma_G e^{-2j\beta d} = -\Gamma_{\rm th}$. Then, we find:

$$\boxed{V_{\rm th} = \frac{V_G Z_0}{Z_G + Z_0}\,\frac{2e^{-j\beta d}}{1-\Gamma_{\rm th}}\,, \quad I_{\rm sc} = \frac{V_G}{Z_G + Z_0}\,\frac{2e^{-j\beta d}}{1+\Gamma_{\rm th}}} \tag{10.11.4}$$

It follows that $V_{\rm th}/I_{\rm sc} = Z_{\rm th}$, as given by Eq. (10.11.3). A more convenient way of writing Eq. (10.11.4) is by noting the relationships:

$$1 - \Gamma_{\rm th} = \frac{2Z_0}{Z_{\rm th} + Z_0}\,, \quad 1 + \Gamma_{\rm th} = \frac{2Z_{\rm th}}{Z_{\rm th} + Z_0}$$

Then, Eq. (10.11.4) becomes:

$$V_{\rm th} = V_0\,\frac{Z_{\rm th} + Z_0}{Z_0}\,, \quad I_{\rm sc} = I_0\,\frac{Z_{\rm th} + Z_0}{Z_{\rm th}} \tag{10.11.5}$$

where V_0, I_0 are the load voltage and currents in the matched case, given by Eq. (10.9.9). The intuitive meaning of these expressions can be understood by writing them as:

$$V_0 = V_{\rm th}\,\frac{Z_0}{Z_{\rm th} + Z_0}\,, \quad I_0 = I_{\rm sc}\,\frac{Z_{\rm th}}{Z_{\rm th} + Z_0} \tag{10.11.6}$$

These are recognized to be the ordinary voltage and current dividers obtained by connecting the Thévenin and Norton equivalent circuits to the matched load impedance Z_0, as shown in Fig. 10.11.2.

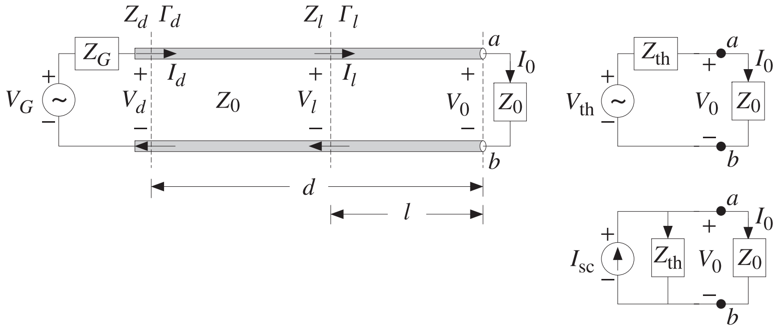


Fig. 10.11.2 Thévenin and Norton equivalent circuits connected to a matched load.

The quantities V_0, I_0 are the same as those obtained by connecting the actual line to the matched load, as was done in Eq. (10.9.9).

An alternative way of determining the quantities $V_{\rm th}$ and $Z_{\rm th}$ is by replacing the length-d transmission line segment by its T-section equivalent circuit, as shown in Fig. 10.11.3.

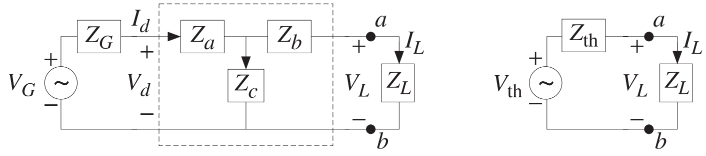


Fig. 10.11.3 T-section and Thévenin equivalent circuits.

The Thévenin equivalent circuit to the left of the terminals a, b is easily determined by shorting the generator and finding the Thévenin impedance and then finding the open-circuit voltage. We have:

$$Z_{\text{th}} = Z_b + \frac{Z_c(Z_a + Z_G)}{Z_c + Z_a + Z_G}\,, \quad V_{\text{th}} = \frac{V_G Z_c}{Z_c + Z_a + Z_G} \tag{10.11.7}$$

where Z_a, Z_b, Z_c for a length-d segment are given by Eq. (10.8.3):

$$Z_a = Z_b = jZ_0 \tan\left(\frac{\beta d}{2}\right)\,, \quad Z_c = -jZ_0\,\frac{1}{\sin\beta d}$$

It is straightforward to verify that the expressions in Eq. (10.11.7) are equivalent to those in Eq. (10.11.3) and (10.11.4).

Example 10.11.1: For the generator, line, and load of Example 10.10.1, determine the Thévenin equivalent circuit. Using this circuit determine the load voltage.

Solution: We work with the T-section approach. The following MATLAB call gives Z_a and Z_c, with $Z_0 = 50$ and $\beta d = 9.5756$:

$$[Z_a, Z_c] = \text{tsection}(50,\ 9.5756) = [-661.89j,\ 332.83j]$$

Then, Eq. (10.11.7) gives with $Z_b = Z_a$:

$$\begin{aligned} Z_{\text{th}} &= Z_b + \frac{Z_c(Z_a + Z_G)}{Z_c + Z_a + Z_G} = 20.39 + j6.36\ \Omega \\ V_{\text{th}} &= \frac{V_G Z_c}{Z_c + Z_a + Z_G} = -10.08 + j0.61 = 10.10\angle 176.52^{\text{o}}\ \text{V} \end{aligned}$$

Alternatively, Z_{th} can be computed by propagating $Z_G = 20$ by a distance d:

$$Z_{\text{th}} = \text{zprop}(20, 50, 9.5756) = 20.39 + j6.36\ \Omega$$

The load voltage is found from the Thévenin circuit:

$$V_L = \frac{V_{\text{th}} Z_L}{Z_L + Z_{\text{th}}} = -7.09 + j0.65 = 7.12\angle 174.75^{\text{o}}\ \text{V}$$

which agrees with that found in Example 10.10.1. □

10.12 Standing Wave Ratio

The line voltage at a distance l from the load is given by Eq. (10.9.12), which can be written as follows in terms of the forward wave $V_{L+} = V_L/(1 + \Gamma_L)$:

$$V_l = V_{L+} e^{j\beta l}(1 + \Gamma_l) \tag{10.12.1}$$

The magnitude of V_l will be:

$$|V_l| = |V_{L+}||1 + \Gamma_l| = |V_{L+}||1 + \Gamma_L e^{-2j\beta l}| \tag{10.12.2}$$

It follows that $|V_l|$ will vary sinusoidally as a function of l. Its limits of variation are determined by noting that the quantity $|1 + \Gamma_l|$ varies between:

$$1 - |\Gamma_L| = 1 - |\Gamma_l| \le |1 + \Gamma_l| \le 1 + |\Gamma_l| = 1 + |\Gamma_L|$$

where we used $|\Gamma_l| = |\Gamma_L|$. Thus, $|V_l|$ will vary over the limits:

$$V_{\min} \le |V_l| \le V_{\max} \tag{10.12.3}$$

where

$$\begin{aligned} V_{\min} &= |V_{L+}| - |V_{L-}| = |V_{L+}|\big(1 - |\Gamma_L|\big) \\ V_{\max} &= |V_{L+}| + |V_{L-}| = |V_{L+}|\big(1 + |\Gamma_L|\big) \end{aligned} \tag{10.12.4}$$

We note that the reflection coefficient at a load $Z_L = R_L + jX_L$ has always magnitude less than unity, $|\Gamma_L| \le 1$. Indeed, this follows from the positivity of R_L and the following property:

$$Z_L = Z_0\,\frac{1 + \Gamma_L}{1 - \Gamma_L} \quad \Rightarrow \quad R_L = \text{Re}(Z_L) = Z_0\,\frac{1 - |\Gamma_L|^2}{|1 + \Gamma_L|^2} \tag{10.12.5}$$

The voltage *standing wave ratio* (SWR) of a terminated transmission line is a measure of the degree of matching of the line to the load and is defined as the ratio of the maximum to minimum voltage along the line:

$$\boxed{S = \frac{V_{\max}}{V_{\min}} = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|}} \quad \Leftrightarrow \quad \boxed{|\Gamma_L| = \frac{S - 1}{S + 1}} \tag{10.12.6}$$

Because $|\Gamma_L| \le 1$, the SWR will always be $S \ge 1$. A matched load, $\Gamma_L = 0$, has $S = 1$. The more unmatched the load is, the larger the SWR. Indeed, $S \to \infty$ as $|\Gamma_L| \to 1$. A matched line has $V_{\min} = |V_l| = V_{\max}$ at all points l, and is sometimes referred to as a *flat line.* The MATLAB function `swr.m` calculates the SWR from Eq. (10.12.6):

```
S = swr(Gamma);                 % calculates SWR from reflection coefficient Γ
```

The SWR can be used to quantify the amount of power delivered to the load. The percentage of reflected power from the load is $|\Gamma_L|^2$. Therefore, the percentage of the power delivered to the load relative to the incident power will be:

$$\frac{P_L}{P_{\text{inc}}} = 1 - |\Gamma_L|^2 = \frac{4S}{(S + 1)^2} \tag{10.12.7}$$

The larger the SWR, the smaller the percentage of delivered power. For example, if $S = 9$, the reflection coefficient will have magnitude $|\Gamma_L| = 0.8$, resulting in $1 - |\Gamma_L|^2 = 0.36$, that is, only 36 percent of the incident power gets transferred to the load.

Example 10.12.1: If the reflected wave at the load of a transmission line is 6 dB below the incident wave, what is the SWR at the load? What percentage of the incident power gets transferred to the load?

Solution: The relative power levels of the reflected and incident waves will be:

$$|\Gamma_L|^2 = \frac{|V_-|^2}{|V_+|^2} = 10^{-6/10} = \frac{1}{4} \quad\Rightarrow\quad |\Gamma_L| = \frac{1}{2} \quad\Rightarrow\quad S = \frac{1+0.5}{1-0.5} = 3$$

The fraction of power transferred to the load is $1 - |\Gamma_L|^2 = 0.75$, or 75 percent. □

If both the line and load impedances are real-valued, then the standing wave ratio is $S = Z_L/Z_0$ if $Z_L \geq Z_0$, and $S = Z_0/Z_L$, if $Z_L \leq Z_0$. This follows from the identity:

$$\boxed{S = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = \frac{|Z_L+Z_0| + |Z_L-Z_0|}{|Z_L+Z_0| - |Z_L-Z_0|} = \frac{\max(Z_L, Z_0)}{\min(Z_L, Z_0)}} \tag{10.12.8}$$

or, explicitly:

$$S = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = \begin{cases} \dfrac{Z_L}{Z_0}\,, & \text{if } Z_L \geq Z_0 \\[2ex] \dfrac{Z_0}{Z_L}\,, & \text{if } Z_L \leq Z_0 \end{cases} \tag{10.12.9}$$

10.13 Determining an Unknown Load Impedance

Often a transmission line is connected to an unknown impedance, and we wish to determine that impedance by making appropriate measurements of the voltage along the line.

The SWR can be readily determined by measuring $|V_l|$ and finding its maximum and minimum values V_{max} and V_{min}. From the SWR, we then determine the magnitude of the reflection coefficient $|\Gamma_L|$.

The phase of Γ_L can be determined by finding the locations along the line at which a voltage maximum or a voltage minimum is measured. If θ_L is the required phase, so that $\Gamma_L = |\Gamma_L|e^{j\theta_L}$, then we have:

$$|V_l| = |V_{L+}||1+\Gamma_l| = |V_{L+}||1+\Gamma_L e^{-2j\beta l}| = |V_{L+}|\,\big|1+|\Gamma_L|e^{j(\theta_L - 2\beta l)}\,\big|$$

At all locations l for which $\theta_L - 2\beta l = \pm 2\pi n$, where n is an integer, we will have $\Gamma_l = |\Gamma_L|$ and $|V_l|$ will be equal to V_{max}. Similarly, at all locations for which $\theta_L - 2\beta l = \pm(2n+1)\pi$, we will have $\Gamma_l = -|\Gamma_L|$ and $|V_l|$ will be equal to V_{min}.

We note that two successive maxima, or two successive minima, are separated by a distance $\lambda/2$ and a maximum is separated by the next minimum by a distance $\lambda/4$, so that $|l_{\text{max}} - l_{\text{min}}| = \lambda/4$.

Once such distances $l_{\text{max}}, l_{\text{min}}$ have been determined, the full reflection coefficient can be constructed from $\Gamma_L = \Gamma_l e^{2j\beta l}$, where $\Gamma_l = \pm|\Gamma_L|$ depending on using a maximum- or minimum-voltage distance l. From Γ_L and the knowledge of the line impedance Z_0, the load impedance Z_L can be computed. Thus, we have:

$$\boxed{\Gamma_L = |\Gamma_L|e^{j\theta_L} = |\Gamma_L|e^{2j\beta l_{\text{max}}} = -|\Gamma_L|e^{2j\beta l_{\text{min}}} \quad\Rightarrow\quad Z_L = Z_0\frac{1+\Gamma_L}{1-\Gamma_L}} \tag{10.13.1}$$

If $0 \leq \theta_L \leq \pi$, the locations for the closest maxima and minima to the load are determined from the conditions:

$$\theta_L - 2\beta l_{\text{max}} = 0\,, \quad \theta_L - 2\beta l_{\text{min}} = -\pi$$

resulting in the distances:

$$\boxed{l_{\text{max}} = \frac{\theta_L}{4\pi}\lambda\,, \quad l_{\text{min}} = \frac{\theta_L + \pi}{4\pi}\lambda}\,, \quad (0 \leq \theta_L \leq \pi) \tag{10.13.2}$$

Similarly, if $-\pi \leq \theta_L \leq 0$, we must solve $\theta_L - 2\beta l_{\text{max}} = -2\pi$ and $\theta_L - 2\beta l_{\text{min}} = -\pi$:

$$\boxed{l_{\text{max}} = \frac{\theta_L + 2\pi}{4\pi}\lambda\,, \quad l_{\text{min}} = \frac{\theta_L + \pi}{4\pi}\lambda}\,, \quad (-\pi \leq \theta_L \leq 0) \tag{10.13.3}$$

Of course, one wants to solve for θ_L in terms of the measured l_{max} or l_{min}. Using l_{min} is more convenient than using l_{max} because θ_L is given by the same expression in both cases. The lengths $l_{\text{max}}, l_{\text{min}}$ may be assumed to be less than $\lambda/2$ (if not, we may subtract enough multiples of $\lambda/2$ until they are.) Expressing θ_L in terms of the measured l_{min}, we have:

$$\boxed{\theta_L = \frac{4\pi l_{\text{min}}}{\lambda} - \pi = 2\beta l_{\text{min}} - \pi} \tag{10.13.4}$$

Alternatively, we have in terms of l_{max}:

$$\boxed{\theta_L = \begin{cases} \dfrac{4\pi l_{\text{max}}}{\lambda} = 2\beta l_{\text{max}} & \text{if} \quad 0 \leq l_{\text{max}} \leq \dfrac{\lambda}{4} \\[2ex] \dfrac{4\pi l_{\text{max}}}{\lambda} - 2\pi = 2\beta l_{\text{max}} - 2\pi & \text{if} \quad \dfrac{\lambda}{4} \leq l_{\text{max}} \leq \dfrac{\lambda}{2} \end{cases}} \tag{10.13.5}$$

Example 10.13.1: A 50-ohm line is connected to an unknown impedance. Voltage measurements along the line reveal that the maximum and minimum voltage values are 1.75 and 0.25 volts, respectively. Moreover, the closest distance to the load at which a voltage maximum is observed is 0.125λ.

Determine the reflection coefficient Γ_L, the load impedance Z_L, and the closest distance to the load at which a voltage minimum is observed.

For another load, the same maxima and minima are observed, but now the closest distance to the load at which a minimum is observed is 0.125λ. Determine Γ_L and Z_L.

Solution: The SWR is determined to be $S = V_{\text{max}}/V_{\text{min}} = 1.75/0.25 = 7$. Then, the magnitude of the reflection coefficient is found to be $|\Gamma_L| = (S-1)/(S+1) = (7-1)/(7+1) = 0.75$.

Given that at $l_{\text{max}} = \lambda/8$ we observe a voltage maximum, we compute the phase from Eq. (10.13.5), $\theta_L = 2\beta l_{\text{max}} = 4\pi/8 = \pi/2$. Then, the reflection coefficient will be:

$$\Gamma_L = |\Gamma_L|e^{j\theta_L} = 0.75e^{j\pi/2} = 0.75j$$

It follows that the load impedance will be:

$$Z_L = Z_0\,\frac{1+\Gamma_L}{1-\Gamma_L} = 50\,\frac{1+0.75j}{1-0.75j} = 14+48j\ \Omega$$

The closest voltage minimum will occur at $l_{\min} = l_{\max} + \lambda/4 = 0.375\lambda = 3\lambda/8$. Alternatively, we could have determined the phase from Eq. (10.13.4), $\theta_L = 2\beta l_{\min} - \pi = 4\pi(3/8) - \pi = \pi/2$. The left graph of Fig. 10.13.1 shows a plot of $|V_l|$ versus l.

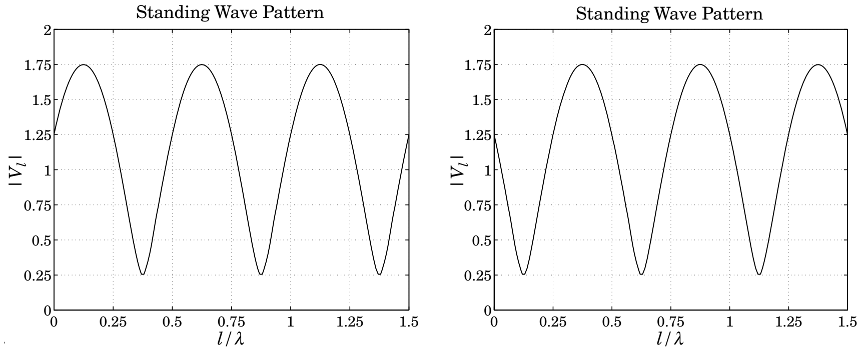


Fig. 10.13.1 Standing wave patterns.

Note the locations of the closest voltage maxima and minima to the load, that is $\lambda/8$ and $3\lambda/8$. In the second case, we are given $l_{\min} = \lambda/8$. It follows that $\theta_L = 2\beta l_{\min} - \pi = \pi/2 - \pi = -\pi/2$. Alternatively, we may work with $l_{\max} = l_{\min} + \lambda/4 = 3\lambda/8$. Because $l_{\max} > \lambda/4$, Eq. (10.13.5) will give $\theta_L = 2\beta l_{\max} - 2\pi = 4\pi(3/8) - 2\pi = -\pi/2$. The reflection coefficient and load impedance will be:

$$\Gamma_L = |\Gamma_L|e^{j\theta_L} = 0.75e^{-j\pi/2} = -0.75j \quad\Rightarrow\quad Z_L = 14-48j\ \Omega$$

The right graph of Fig. 10.13.1 depicts the standing wave pattern in this case. □

It is interesting also to determine the wave impedances at the locations along the line at which we have voltage maxima or minima, that is, at $l = l_{\max}$ or $l_{\min}$. The answers are expressed in terms of the SWR. Indeed, at $l = l_{\max}$, we have $\Gamma_l = |\Gamma_L|$ which gives:

$$Z_{\max} = Z_0\,\frac{1+\Gamma_l}{1-\Gamma_l} = Z_0\,\frac{1+|\Gamma_L|}{1-|\Gamma_L|} = S\,Z_0 \tag{10.13.6}$$

Similarly, at $l = l_{\min}$, we have $\Gamma_l = -|\Gamma_L|$ and find:

$$Z_{\min} = Z_0\,\frac{1+\Gamma_l}{1-\Gamma_l} = Z_0\,\frac{1-|\Gamma_L|}{1+|\Gamma_L|} = \frac{1}{S}\,Z_0 \tag{10.13.7}$$

We note that $Z_{\max}Z_{\min} = Z_0^2$, as is expected because the points $l_{\max}$ and $l_{\min}$ are separated by a quarter-wavelength distance $\lambda/4$.

Because at $l_{\max}$ and $l_{\min}$ the wave impedances are *real-valued*, these points can be used as convenient locations at which to insert a quarter-wave transformer to match a line with real Z_0 to a complex load Z_L. Given θ_L, the required locations are determined from Eq. (10.13.2) or (10.13.3). We discuss this matching method later on.

The MATLAB function `lmin.m` calculates the locations $l_{\min}$ and $l_{\max}$ from Eqs. (10.13.2) and (10.13.3), and the corresponding impedances $Z_{\min}$ and $Z_{\max}$. It has usage:

```
[lm,Zm] = lmin(ZL,Z0,'min');          % locations of voltage minima
[lm,Zm] = lmin(ZL,Z0,'max');          % locations of voltage maxima
```

For a lossless line the power delivered to the load can be measured at any point l along the line, and in particular, at $l_{\max}$ and $l_{\min}$. Then, Eq. (10.12.7) can be written in the alternative forms:

$$P_L = \frac{1}{2Z_0}\left(|V_{L+}|^2 - |V_{L-}|^2\right) = \frac{V_{\max}V_{\min}}{2Z_0} = \frac{V_{\min}^2}{2Z_{\min}} = \frac{V_{\max}^2}{2Z_{\max}} = \frac{V_{\max}^2}{2SZ_0} \tag{10.13.8}$$

The last expression shows that for a given maximum voltage that can be supported along a line, the power transmitted to the load is S times smaller than it could be if the load were matched.

Conversely, for a given amount P_L of transmitted power, the maximum voltage will be $V_{\max} = \sqrt{2SP_LZ_0}$. One must ensure that for a highly unmatched load, $V_{\max}$ remain less than the breakdown voltage of the line.

If the line is lossy, measurements of the SWR along its length will give misleading results. Because the reflected power attenuates as it propagates backwards away from the load, the SWR will be smaller at the line input than at the load.

For a lossy line with $\beta_c = \beta - j\alpha$, the reflection coefficient at the line input will be: $\Gamma_d = \Gamma_L e^{-2(\alpha+j\beta)d}$, which gives for the input SWR:

$$S_d = \frac{1+|\Gamma_d|}{1-|\Gamma_d|} = \frac{1+|\Gamma_L|e^{-2\alpha d}}{1-|\Gamma_L|e^{-2\alpha d}} = \frac{e^{2\alpha d}+|\Gamma_L|}{e^{2\alpha d}-|\Gamma_L|} = \frac{a+|\Gamma_L|}{a-|\Gamma_L|} \tag{10.13.9}$$

where we expressed it in terms of the matched-line loss of Eq. (10.10.7).

Example 10.13.2: For the RG-58 coax cable of Example 10.10.2, we find the SWRs:

$$S_L = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = \frac{1+0.62}{1-0.62} = 4.26\,, \quad S_d = \frac{1+|\Gamma_d|}{1-|\Gamma_d|} = \frac{1+0.41}{1-0.41} = 2.39$$

If one does not know that the line is lossy, and measures the SWR at the line input, one would think that the load is more matched than it actually is. □

Example 10.13.3: The SWR at the load of a line is 9. If the matched-line loss is 10 dB, what is the SWR at the line input?

Solution: We calculate the reflection coefficient at the load:

$$|\Gamma_L| = \frac{S-1}{S+1} = \frac{9-1}{9+1} = 0.8$$

The matched-line loss is $a = 10^{L_M/10} = 10^{10/10} = 10$. Thus, the reflection coefficient at the input will be $|\Gamma_d| = |\Gamma_L|/a = 0.8/10 = 0.08$. The corresponding SWR will be $S = (1+0.08)/(1-0.08) = 1.17$. □

Example 10.13.4: A 50-ohm line feeds a half-wave dipole antenna with impedance of $73 + j42.5$ ohms. The line has matched-line loss of 3 dB. What is the total loss of the line? What is the SWR at the load and at the line input?

If the line length is doubled, what is the matched-line loss, the total loss, the input and load SWRs?

Solution: The matched-line loss in absolute units is $a = 10^{3/10} = 2$. Using the MATLAB functions `z2g` and `swr`, we compute the reflection coefficient at the load and its SWR:

$$|\Gamma_L| = \left|\frac{Z_L - Z_0}{Z_L + Z_0}\right| = \left|\frac{73 + j42.5 - 50}{73 + j42.5 + 50}\right| = \text{abs}\big(\text{z2g}(73 + 42.5j, 50)\big) = 0.3713$$

The SWR will be $S = \text{swr}(0.3713) = 2.1814$. The reflection coefficient at the line input will be $|\Gamma_d| = |\Gamma_L|e^{-2\alpha d} = |\Gamma_L|/a = 0.1857$, and its SWR, $S = \text{swr}(0.1857) = 1.4560$.

If the line length is doubled, the matched-line loss in dB will double to 6 dB, since it is given by $L_M = 8.686\alpha d$. In absolute units, it is $a = 2^2 = 4$.

The corresponding reflection coefficient at the line input will be $|\Gamma_d| = |\Gamma_L|/a = 0.0928$, and its SWR, $S = \text{swr}(0.0928) = 1.2047$. □

10.14 Smith Chart

The relationship between the wave impedance Z and the corresponding reflection response Γ along a transmission line Z_0 can be stated in terms the normalized impedance $z = Z/Z_0$ as follows:

$$\Gamma = \frac{z-1}{z+1} \quad \Leftrightarrow \quad z = \frac{1+\Gamma}{1-\Gamma} \tag{10.14.1}$$

It represents a mapping between the complex impedance z-plane and the complex reflection coefficient Γ-plane, as shown in Fig. 10.14.1. The mapping is similar to the bilinear transformation mapping in linear system theory between the s-plane (playing the role of the impedance plane) and the z-plane of the z-transform (playing the role of the Γ-plane.)

A complex impedance $z = r + jx$ with positive resistive part, $r > 0$, gets mapped onto a point Γ that lies inside the unit-circle in the Γ-plane, that is, satisfying $|\Gamma| < 1$.

An entire resistance line $z = r$ (a vertical line on the z-plane) gets mapped onto a circle on the Γ-plane that lies entirely inside the unit-circle, if $r > 0$. Similarly, a reactance line $z = jx$ (a horizontal line on the z-plane) gets mapped onto a circle on the Γ-plane, a portion of which lies inside the unit-circle.

The Smith chart is a graphical representation of the Γ-plane with a curvilinear grid of constant resistance and constant reactance circles drawn inside the unit-circle. In effect, the Smith chart is a curvilinear graph paper.

Any reflection coefficient point Γ falls at the intersection of a resistance and a reactance circle, r, x, from which the corresponding impedance can be read off immediately as $z = r + jx$. Conversely, given $z = r + jx$ and finding the intersection of the r, x circles, the complex point Γ can be located and its value read off in polar or cartesian coordinates.

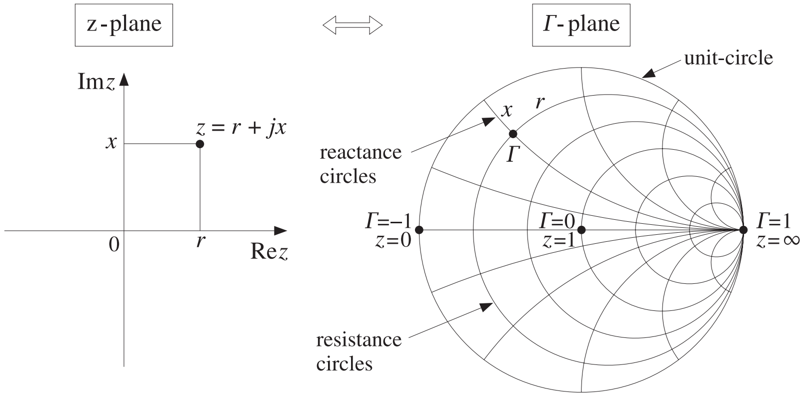


Fig. 10.14.1 Mapping between z-plane and Γ-plane.

To determine the centers and radii of the resistance and reactance circles, we use the result that a circle with center C and radius R on the Γ-plane has the following two equivalent representations:

$$|\Gamma|^2 - C^*\Gamma - C\Gamma^* = B \quad \Leftrightarrow \quad |\Gamma - C| = R\,, \qquad \text{where } B = R^2 - |C|^2 \tag{10.14.2}$$

Setting $z = r + jx$ in Eq. (10.14.1) and extracting the real and imaginary parts, we can write r and x in terms of Γ, as follows:

$$r = \text{Re}\, z = \frac{1 - |\Gamma|^2}{|1-\Gamma|^2}\,, \quad x = \text{Im}\, z = \frac{j(\Gamma^* - \Gamma)}{|1-\Gamma|^2} \tag{10.14.3}$$

In particular, the expression for the resistive part implies that the condition $r > 0$ is equivalent to $|\Gamma| < 1$. The r, x circles are obtained by putting Eqs. (10.14.3) in the form of Eq. (10.14.2). We have:

$$r|\Gamma - 1|^2 = 1 - |\Gamma|^2 \quad \Rightarrow \quad r\big(|\Gamma|^2 - \Gamma - \Gamma^* + 1\big) = 1 - |\Gamma|^2$$

and rearranging terms:

$$|\Gamma|^2 - \frac{r}{r+1}\Gamma - \frac{r}{1+r}\Gamma^* = \frac{1-r}{1+r} \quad \Rightarrow \quad \left|\Gamma - \frac{r}{1+r}\right|^2 = \frac{1-r}{1+r} + \frac{r^2}{(1+r)^2} = \left(\frac{1}{1+r}\right)^2$$

Similarly, we have

$$x|\Gamma - 1|^2 = j(\Gamma^* - \Gamma) \quad \Rightarrow \quad x\big(|\Gamma|^2 - \Gamma - \Gamma^* + 1\big) = j(\Gamma^* - \Gamma)$$

which can be rearranged as:

$$|\Gamma|^2 - \left(1 - \frac{j}{x}\right)\Gamma - \left(1 + \frac{j}{x}\right)\Gamma^* = -1 \quad \Rightarrow \quad \left|\Gamma - \left(1 + \frac{j}{x}\right)\right|^2 = -1 + \left(1 + \frac{1}{x^2}\right) = \left(\frac{1}{x}\right)^2$$

To summarize, the constant resistance and reactance circles are:

$$
\begin{aligned}
\left|\Gamma - \frac{r}{1+r}\right| &= \frac{1}{1+r} \qquad \text{(resistance circles)}\\
\left|\Gamma - \left(1+\frac{j}{x}\right)\right| &= \frac{1}{|x|} \qquad \text{(reactance circles)}
\end{aligned}
\tag{10.14.4}
$$

The centers of the resistance circles are on the positive half of the real axis on the Γ-plane, lying between $0 \le \Gamma \le 1$. When $r = 0$, the impedance circle is the entire unit-circle with center at $\Gamma = 0$. As r increases, the radii become smaller and the centers move towards $\Gamma = 1$. The centers of the reactance circles lie on the tangent of the unit-circle at $\Gamma = 1$.

Example 10.14.1: Fig. 10.14.2 depicts the resistance and reactance circles for the following values of r, x:

$$r = [0.2,\ 0.5,\ 1,\ 2,\ 5]\,,\quad x = [0.2,\ 0.5,\ 1,\ 2,\ 5]$$

Because the point A is at the intersection of the $r = 0.2$ and $x = 0.5$ circles, the corresponding impedance will be $z_A = 0.2 + 0.5j$. We list below the impedances and reflection coefficients at the points A, B, C, D, E, S, P, O:

$$
\begin{array}{lll}
 & z_A = 0.2 + 0.5j, & \Gamma_A = -0.420 + 0.592j = 0.726\angle 125.37^{\rm o}\\
 & z_B = 0.5 - j, & \Gamma_B = 0.077 - 0.615j = 0.620\angle -82.88^{\rm o}\\
 & z_C = 2 - 2j, & \Gamma_C = 0.539 - 0.308j = 0.620\angle -29.74^{\rm o}\\
 & z_D = j, & \Gamma_D = j = 1\angle 90^{\rm o}\\
 & z_E = -j, & \Gamma_E = -j = 1\angle -90^{\rm o}\\
\text{(short circuit)} & z_S = 0, & \Gamma_S = -1 = 1\angle 180^{\rm o}\\
\text{(open circuit)} & z_P = \infty, & \Gamma_P = 1 = 1\angle 0^{\rm o}\\
\text{(matched)} & z_O = 1, & \Gamma_O = 0 = 0\angle 0^{\rm o}
\end{array}
$$

The points S and P correspond to a short-circuited and an open-circuited impedance. The center of the Smith chart at point O corresponds to $z = 1$, that is, an impedance matched to the line. □

The Smith chart helps one visualize the wave impedance as one moves away from or towards a load. Assuming a lossless line, the wave impedance and corresponding reflection response at a distance l from the load are given by:

$$z_l = \frac{z_L + j\tan\beta l}{1 + jz_L\tan\beta l} \quad\Leftrightarrow\quad \Gamma_l = e^{-2j\beta l}\Gamma_L \tag{10.14.5}$$

The magnitude of Γ_l remains constant as l varies, indeed, $|\Gamma_l| = |\Gamma_L|$. On the Smith chart, this represents a circle centered at the origin $\Gamma = 0$ of radius $|\Gamma_L|$. Such circles are called *constant SWR circles* because the SWR is related to the circle radius by

$$S = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|}$$

The relative phase angle between Γ_l and Γ_L is negative, $-2\beta l$, and therefore, the point Γ_l moves clockwise along the constant SWR circle, as shown in Fig. 10.14.3. Conversely, if l is decreasing towards the load, the point Γ_l will be moving counter-clockwise.

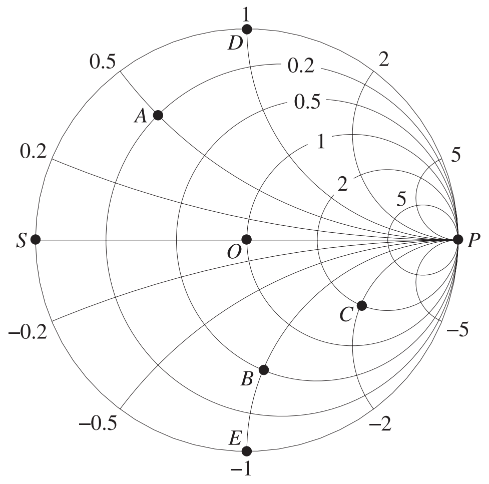


Fig. 10.14.2 Smith chart example.

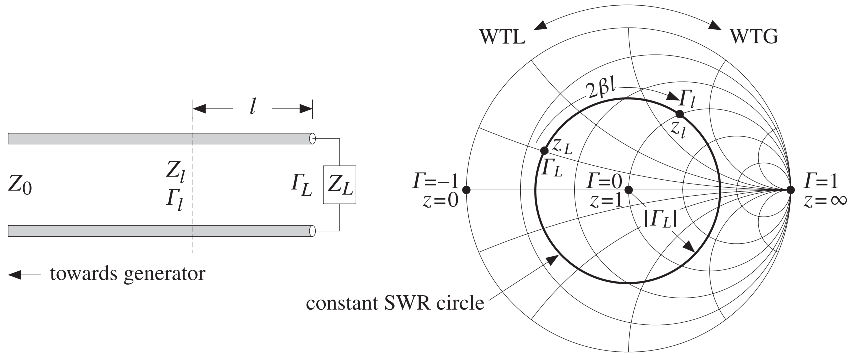


Fig. 10.14.3 Moving towards the generator along a constant SWR circle.

The rotation angle $\phi_l = 2\beta l$ can be read off in degrees from the outer periphery of the Smith chart. The corresponding length l can also be read off in units of wavelengths towards the generator (WTG) or wavelengths towards the load (WTL). Moving towards the generator by a distance $l = \lambda/8$ corresponds to a clockwise rotation by an angle of $\phi_l = 2(2\pi/8) = \pi/2$, that is, $90^{\rm o}$. Moving by $l = \lambda/4$ corresponds to a $180^{\rm o}$ rotation, and by $l = \lambda/2$, to a full $360^{\rm o}$ rotation.

Smith charts provide an intuitive geometrical representation of a load in terms of its reflection coefficient and help one design matching circuits—where matching means moving towards the center of the chart. However, the computational accuracy of the Smith chart is not very high, about 5–10%, because one must visually interpolate between the grid circles of the chart.

Smith charts are used widely to display S-parameters of microwave amplifiers and help with the design of matching circuits. Some of the tools used in such designs are the stability circles, gain circles, and noise figure circles of an amplifier, which are intuitively represented on a Smith chart. We discuss them in Chap. 13.

Various resources, including a history of the Smith chart and high-quality downloadable charts in Postscript format can be found on the web site [1317].

Laursen's Smith chart MATLAB toolbox can be used to draw Smith charts. It is available from the Mathworks web site [1328]. Our MATLAB function `smith.m` can be used to draw simple Smith charts.

10.15 Time-Domain Response of Transmission Lines

So far we discussed only the sinusoidal response of transmission lines. The response to arbitrary time-domain inputs can be obtained by writing Eq. (10.6.3) in the time domain by replacing $j\omega \to \partial/\partial t$. We will assume a lossless line and set $R' = G' = 0$.[†] We obtain then the system of coupled equations:

$$\frac{\partial V}{\partial z} = -L'\frac{\partial I}{\partial t}\,, \qquad \frac{\partial I}{\partial z} = -C'\frac{\partial V}{\partial t} \tag{10.15.1}$$

These are called *telegrapher's equations*. By differentiating again with respect to z, it is easily verified that V and I satisfy the uncoupled one-dimensional wave equations:

$$\frac{\partial^2 V}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0\,, \qquad \frac{\partial^2 I}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 I}{\partial t^2} = 0$$

where $c = 1/\sqrt{L'C'}$. As in Sec. 2.1, it is better to deal directly with the first-order coupled system (10.15.1). This system can be uncoupled by defining the forward and backward wave components:

$$V_\pm(t,z) = \frac{V(t,z) \pm Z_0 I(t,z)}{2}\,, \quad \text{where} \quad Z_0 = \sqrt{\frac{L'}{C'}} \tag{10.15.2}$$

These satisfy the uncoupled equations:

$$\frac{\partial V_\pm}{\partial z} = \mp\frac{1}{c}\frac{\partial V_\pm}{\partial t} \tag{10.15.3}$$

with general solutions given in terms of two arbitrary functions $f(t), g(t)$:

$$V_+(t,z) = f(t - z/c)\,, \quad V_-(t,z) = g(t + z/c) \tag{10.15.4}$$

These solutions satisfy the basic forward and backward propagation property:

$$\boxed{\begin{aligned} V_+(t, z+\Delta z) &= V_+(t - \Delta t, z) \\ V_-(t, z+\Delta z) &= V_-(t + \Delta t, z) \end{aligned}}\,, \qquad \text{where} \quad \Delta t = \frac{\Delta z}{c} \tag{10.15.5}$$

[†] At RF, R', G' may be small but cannot be assumed to be frequency-independent, for example, R' depends on the surface impedance R_s, which grows like $f^{1/2}$.

In particular, we have:

$$\boxed{\begin{aligned} V_+(t,z) &= V_+(t - z/c, 0) \\ V_-(t,z) &= V_-(t + z/c, 0) \end{aligned}} \tag{10.15.6}$$

These allow the determination of the line voltages at any point z along the line from the knowledge of the voltages at $z = 0$. Next, we consider a terminated line, shown in Fig. 10.15.1, driven by a generator voltage $V_G(t)$, which is typically turned on at $t = 0$ as indicated by the closing of the switch.

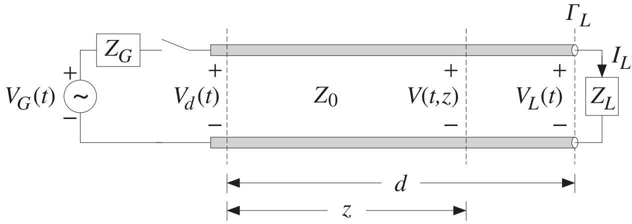


Fig. 10.15.1 Transient response of terminated line.

In general, Z_G and Z_L may have inductive or capacitive parts. To begin with, we will assume that they are purely *resistive*. Let the length of the line be d, so that the one- and two-way travel-time delays will be $T = d/c$ and $2T = 2d/c$.

When the switch closes, an initial waveform is launched forward along the line. When it reaches the load T seconds later, it gets reflected, picking up a factor of Γ_L, and begins to travel backward. It reaches the generator T seconds later, or $2T$ seconds after the initial launch, and gets reflected there traveling forward again, and so on. The total forward- and backward-moving components $V_\pm(t,z)$ include all the multiple reflections.

Before we sum up the multiple reflections, we can express $V_\pm(t,z)$ in terms of the *total* forward-moving component $V_+(t) \equiv V_+(t,0)$ at the generator end, with the help of (10.15.6). In fact, we have $V_+(t,z) = V_+(t - z/c)$. Applying this at the load end $z = d$, we have $V_L^+(t) = V_+(t,d) = V_+(t - d/c) = V_+(t - T)$. Because of Ohm's law at the load, $V_L(t) = Z_L I_L(t)$, we have for the forward/backward components:

$$V_L^\pm(t) = \frac{V_L(t) \pm Z_0 I_L(t)}{2} = \frac{Z_L \pm Z_0}{2}\, I_L(t) \quad\Rightarrow\quad V_L^-(t) = \frac{Z_L - Z_0}{Z_L + Z_0}\, V_L^+(t) = \Gamma_L V_+(t - T)$$

Therefore, we find the total voltage at the load end:

$$\boxed{V_L(t) = V_L^+(t) + V_L^-(t) = (1 + \Gamma_L) V_+(t - T)} \tag{10.15.7}$$

Using (10.15.6), the backward component at $z = 0$ is:

$$\begin{aligned} V_-(t+T) &= V_-(t + d/c, 0) = V_-(t,d) = V_L^-(t) = \Gamma_L V_+(t - T)\,, \quad \text{or,} \\ V_-(t) &= \Gamma_L V_+(t - 2T) \end{aligned}$$

Thus, the total line voltage at the generator end will be:

$$\boxed{V_d(t) = V_+(t) + V_-(t) = V_+(t) + \Gamma_L V_+(t-2T)} \tag{10.15.8}$$

More generally, the voltage at any point z along the line will be:

$$V(t,z) = V_+(t,z) + V_-(t,z) = V_+(t - z/c) + \Gamma_L V_+(t + z/c - 2T) \tag{10.15.9}$$

It remains to determine the total forward component $V_+(t)$ in terms of the multiple reflections of the initially launched wave along the line. We find below that:

$$\begin{aligned} V_+(t) &= \sum_{m=0}^{\infty} (\Gamma_G\Gamma_L)^m V(t - 2mT) \\ &= V(t) + (\Gamma_G\Gamma_L)V(t-2T) + (\Gamma_G\Gamma_L)^2 V(t-4T) + \cdots \end{aligned} \tag{10.15.10}$$

where $V(t)$ is the initially launched waveform:

$$V(t) = \frac{Z_0}{Z_G + Z_0} V_G(t) \tag{10.15.11}$$

Thus, initially the transmission line can be replaced by a voltage divider with Z_0 in series with Z_L. For a right-sided signal $V(t)$, such as that generated after closing the switch, the number of terms in (10.15.10) is finite, but growing with time. Indeed, the requirement that the argument of $V(t - 2mT)$ be non-negative, $t - 2mT \geq 0$, may be solved for the limits on m:

$$0 \leq m \leq M(t)\,, \qquad \text{where} \quad M(t) = \text{floor}\left(\frac{t}{2T}\right) \tag{10.15.12}$$

To justify (10.15.10) and (10.15.11), we may start with the single-frequency case discussed in Sec. 10.9 and perform an inverse Fourier transform. Defining the z-transform variable $\zeta = e^{j\omega T} = e^{j\beta d}$,[†] we may rewrite Eq. (10.9.7) in the form:

$$V_d = V\,\frac{1 + \Gamma_L\zeta^{-2}}{1 - \Gamma_G\Gamma_L\zeta^{-2}}\,, \quad Z_0 I_d = V\,\frac{1 - \Gamma_L\zeta^{-2}}{1 - \Gamma_G\Gamma_L\zeta^{-2}}\,, \quad \text{where} \quad V = \frac{V_G Z_0}{Z_G + Z_0}$$

The forward and backward waves at $z = 0$ will be:

$$\begin{aligned} V_+ &= \frac{V_d + Z_0 I_d}{2} = \frac{V}{1 - \Gamma_G\Gamma_L\zeta^{-2}} \\ V_- &= \frac{V_d - Z_0 I_d}{2} = \frac{V\Gamma_L\zeta^{-2}}{1 - \Gamma_G\Gamma_L\zeta^{-2}} = \Gamma_L\zeta^{-2}V_+ \\ V_d &= V_+ + V_- = V_+ + \Gamma_L\zeta^{-2}V_+ \;\Rightarrow\; V_d(\omega) = V_+(\omega) + \Gamma_L e^{-2j\omega T}V_+(\omega) \end{aligned} \tag{10.15.13}$$

where in the last equation we indicated explicitly the dependence on ω. Using the delay theorem of Fourier transforms, it follows that the equation for $V_d(\omega)$ is the Fourier transform of (10.15.8). Similarly, we have at the load end:

[†] We use ζ instead of z to avoid confusion with the position variable z.

$$V_L = \frac{V_G Z_0}{Z_G + Z_0}\,\frac{1 + \Gamma_L}{1 - \Gamma_G\Gamma_L\zeta^{-2}}\,\zeta^{-1} = (1 + \Gamma_L)\zeta^{-1}V_+$$

which is recognized as the Fourier transform of Eq. (10.15.7). Next, we expand V_+ using the geometric series noting that $|\Gamma_G\Gamma_L\zeta^{-2}| = |\Gamma_G\Gamma_L| < 1$:

$$V_+ = \frac{V}{1 - \Gamma_G\Gamma_L\zeta^{-2}} = V + (\Gamma_G\Gamma_L)\zeta^{-2}V + (\Gamma_G\Gamma_L)^2\zeta^{-4}V + \cdots \tag{10.15.14}$$

which is equivalent to the Fourier transform of Eq. (10.15.10). The same results can be obtained using a *lattice timing diagram*, shown in Fig. 10.15.2, like that of Fig. 5.6.1.

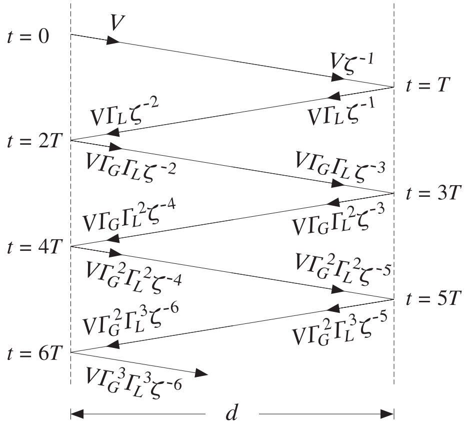

$t = 0$
V
$V\zeta^{-1}$
$t = T$
$V\Gamma_L\zeta^{-2}$
$V\Gamma_L\zeta^{-1}$
$t = 2T$
$V\Gamma_G\Gamma_L\zeta^{-2}$
$V\Gamma_G\Gamma_L\zeta^{-3}$
$t = 3T$
$V\Gamma_G\Gamma_L^2\zeta^{-4}$
$V\Gamma_G\Gamma_L^2\zeta^{-3}$
$t = 4T$
$V\Gamma_G^2\Gamma_L^2\zeta^{-4}$
$V\Gamma_G^2\Gamma_L^2\zeta^{-5}$
$t = 5T$
$V\Gamma_G^2\Gamma_L^3\zeta^{-6}$
$V\Gamma_G^2\Gamma_L^3\zeta^{-5}$
$t = 6T$
$V\Gamma_G^3\Gamma_L^3\zeta^{-6}$
d

Fig. 10.15.2 Lattice timing diagram.

Each propagation segment introduces a delay factor ζ^{-1}, forward or backward, and each reflection at the load and generator ends introduces a factor Γ_L or Γ_G. Summing up all the forward-moving waves at the generator end gives Eq. (10.15.14). Similarly, the summation of the backward terms at the generator, and the summation of the forward and backward terms at the load, give:

$$\begin{aligned} V_- &= V\Gamma_L\zeta^{-2}\big[1 + (\Gamma_G\Gamma_L)\zeta^{-2} + (\Gamma_G\Gamma_L)^2\zeta^{-4} + \cdots\big] = \Gamma_L\zeta^{-2}V_+ \\ V_L^+ &= V\zeta^{-1}\big[1 + (\Gamma_G\Gamma_L)\zeta^{-2} + (\Gamma_G\Gamma_L)^2\zeta^{-4} + \cdots\big] = \zeta^{-1}V_+ \\ V_L^- &= \Gamma_L V\zeta^{-1}\big[1 + (\Gamma_G\Gamma_L)\zeta^{-2} + (\Gamma_G\Gamma_L)^2\zeta^{-4} + \cdots\big] = \Gamma_L\zeta^{-1}V_+ = \Gamma_L V_L^+ \end{aligned}$$

Replacing $V_+(t)$ in terms of (10.15.10), we obtain from (10.15.7) and (10.15.8):

$$\begin{aligned} V_d(t) &= V(t) + \left(1 + \frac{1}{\Gamma_G}\right)\sum_{m=1}^{\infty}(\Gamma_G\Gamma_L)^m V(t - 2mT) \\ V_L(t) &= (1 + \Gamma_L)\sum_{m=0}^{\infty}(\Gamma_G\Gamma_L)^m V\big(t - (2m+1)T\big) \end{aligned} \tag{10.15.15}$$

The line voltage at an arbitrary location z along the line, can be determined from (10.15.9). The substitution of the series expansion of V_+ leads to the expression:

$$V(t,z)=\sum_{m=0}^{\infty}(\Gamma_G\Gamma_L)^m V(t-z/c-2mT)+\Gamma_L\sum_{k=0}^{\infty}(\Gamma_G\Gamma_L)^k V(t+z/c-2kT-2T)$$

For a causal input $V(t)$, the allowed ranges for the summation indices m, k are:

$$0\le m\le \text{floor}\left(\frac{t-z/c}{2T}\right),\quad 0\le k\le \text{floor}\left(\frac{t+z/c-2T}{2T}\right)$$

Example 10.15.1: A terminated line has $Z_0=50$, $Z_G=450$, $Z_L=150\ \Omega$. The corresponding reflection coefficients are calculated to be: $\Gamma_G=0.8$ and $\Gamma_L=0.5$. For simplicity, we take $c=1$, $d=1$, $T=d/c=1$. First, we consider the transient response of the line to a step generator voltage $V_G(t)=10\,u(t)$. The initial voltage input to the line will be: $V(t)=V_G(t)Z_0/(Z_G+Z_0)=10\,u(t)\cdot 50/(450+50)=u(t)$. It follows from (10.15.15) that:

$$V_d(t)=u(t)+2.25\sum_{m=1}^{\infty}(0.4)^m u(t-2mT),\quad V_L(t)=1.5\sum_{m=1}^{\infty}(0.4)^m u\big(t-(2m+1)T\big)$$

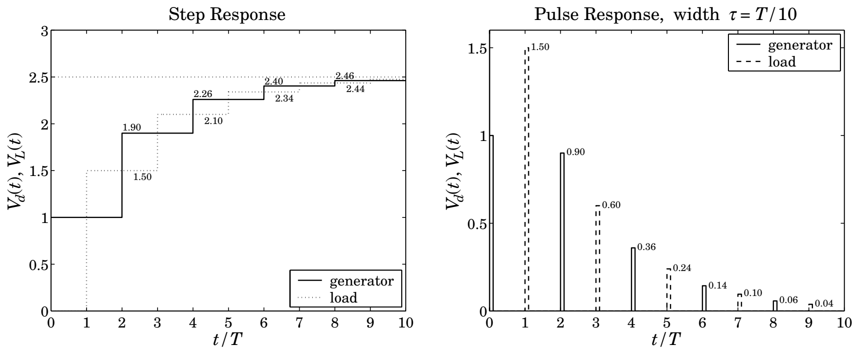


Fig. 10.15.3 Transient step and pulse responses of a terminated line.

These functions are plotted in Fig. 10.15.3. The successive step levels are calculated by:

$V_d(t)$	$V_L(t)$
1	0
$1+2.25[0.4^1]=1.90$	1.5
$1+2.25[0.4^1+0.4^2]=2.26$	$1.5[1+0.4^1]=2.10$
$1+2.25[0.4^1+0.4^2+0.4^3]=2.40$	$1.5([1+0.4^1+0.4^2]=2.34$
$1+2.25[0.4^1+0.4^2+0.4^3+0.4^4]=2.46$	$1.5([1+0.4^1+0.4^2+0.4^3]=2.44$

Both V_d and V_L converge to the same asymptotic value:

$$1+2.25[0.4^1+0.4^2+0.4^3+0.4^4+\cdots]=1.5[1+0.4^1+0.4^2+0.4^3+\cdots]=\frac{1.5}{1-0.4}=2.5$$

More generally, the asymptotic level for a step input $V_G(t)=V_G u(t)$ is found to be:

$$V_\infty=V\frac{1+\Gamma_L}{1-\Gamma_G\Gamma_L}=\frac{V_G Z_0}{Z_G+Z_0}\,\frac{1+\Gamma_L}{1-\Gamma_G\Gamma_L}=\frac{V_G Z_L}{Z_G+Z_L}\tag{10.15.16}$$

Thus, the line behaves asymptotically like a lumped circuit voltage divider with Z_L in series with Z_G. We consider next, the response to a pulse input $V_G(t)=10\big[u(t)-u(t-\tau)\big]$, so that $V(t)=u(t)-u(t-\tau)$, where τ is the pulse duration. Fig. 10.15.3 shows the generator and load line voltages for the case $\tau=T/10=1/10$. The pulse levels are:

$$[1,\ 2.25(0.4)^m]=[1.00,\ 0.90,\ 0.36,\ 0.14,\ 0.06,\ \dots]\quad\text{(at generator)}$$

$$1.5(0.4)^m=[1.50,\ 0.60,\ 0.24,\ 0.10,\ 0.04,\ \dots]\quad\text{(at load)}$$

The following MATLAB code illustrates the computation of $V_d(t)$:

```
d = 1; c=1; T = d/c; tau = T/10; VG = 10;
Z0 = 50; ZG = 450; ZL = 150;
V = VG * Z0 / (ZG+Z0);
gG = z2g(ZG,Z0); gL = z2g(ZL,Z0);          % reflection coefficients ΓG, ΓL

t = 0 : T/1500 : 10*T;

for i=1:length(t),
 M = floor(t(i)/2/T);
 Vd(i) = V * upulse(t(i), tau);
 if M >= 1,
   m = 1:M;
   Vd(i) = Vd(i) + (1+1/gG)*V*sum((gG*gL).^m .* upulse(t(i)-2*m*T, tau));
 end
end

plot(t, Vd, 'r');
```

where *upulse*(t,τ) generates the unit-pulse function $u(t)-u(t-\tau)$. The code can be adapted for any other input function $V(t)$.

The MATLAB file `pulsemovie.m` generates a movie of the step or pulse input as it propagates back and forth between generator and load. It plots the voltage $V(t,z)$ as a function of z at successive time instants t. □

Next, we discuss briefly the case of reactive terminations. These are best handled using Laplace transforms. Introducing the s-domain variable $s=j\omega$, we write $\zeta^{-1}=e^{-j\omega T}=e^{-sT}$. The terminating impedances, and hence the reflection coefficients, become functions of s. For example, if the load is a resistor in series with an inductor, we have $Z_L(s)=R+sL$. Indicating explicitly the dependence on s, we have:

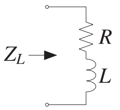
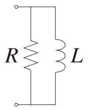
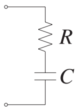
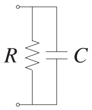
$$V_+(s)=\frac{V(s)}{1-\Gamma_G(s)\Gamma_L(s)e^{-2sT}},\quad\text{where}\quad V(s)=\frac{V_G(s)Z_0}{Z_G(s)+Z_0}\tag{10.15.17}$$

In principle, we may perform an inverse Laplace transform on $V_+(s)$ to find $V_+(t)$. However, this is very tedious and we will illustrate the method only in the case of a matched generator, that is, when $Z_G=Z_0$, or, $\Gamma_G=0$. Then, $V_+(s)=V(s)$, where

$V(s)=V_G(s)Z_0/2Z_0=V_G(s)/2$. The line voltages at the generator and load ends will be from (10.15.13) and (10.15.7):

$$\begin{aligned} V_d(s) &= V(s)+\Gamma_L(s)e^{-2sT}V(s)\\ V_L(s) &= \big[1+\Gamma_L(s)\big]e^{-sT}V(s)\end{aligned} \tag{10.15.18}$$

We consider the four typical cases of series and parallel R–L and series and parallel R–C loads. The corresponding $Z_L(s)$ and $\Gamma_L(s)$ are shown below, where in all cases $\Gamma_R=(R-Z_0)/(R+Z_0)$ and the parameter a gives the effective time constant of the termination, $\tau=1/a$:

series R–L	parallel R–L	series R–C	parallel R–C
			
$Z_L(s)=R+sL$	$Z_L(s)=\dfrac{RsL}{R+sL}$	$Z_L=R+\dfrac{1}{sC}$	$Z_L(s)=\dfrac{R}{1+RCs}$
$\Gamma_L(s)=\dfrac{s+a\Gamma_R}{s+a}$	$\Gamma_L(s)=\dfrac{s\Gamma_R-a}{s+a}$	$\Gamma_L(s)=\dfrac{s\Gamma_R+a}{s+a}$	$\Gamma_L(s)=\dfrac{-s+a\Gamma_R}{s+a}$
$a=\dfrac{R+Z_0}{L}$	$a=\dfrac{Z_0R}{(R+Z_0)L}$	$a=\dfrac{1}{(R+Z_0)C}$	$a=\dfrac{R+Z_0}{RZ_0C}$

We note that in all cases $\Gamma_L(s)$ has the form: $\Gamma_L(s)=(b_0s+b_1)/(s+a)$. Assuming a step-input $V_G(t)=2V_0\,u(t)$, we have $V(t)=V_0\,u(t)$, so that $V(s)=V_0/s$. Then,

$$V_d(s)=V_0\left[\frac{1}{s}+\Gamma_L(s)\frac{1}{s}e^{-2sT}\right]=V_0\left[\frac{1}{s}+\frac{b_0s+b_1}{s(s+a)}e^{-2sT}\right] \tag{10.15.19}$$

Using partial-fraction expansions and the delay theorem of Laplace transforms, we find the inverse Laplace transform:

$$V_d(t)=V_0\,u(t)+V_0\left[\frac{b_1}{a}+\left(b_0-\frac{b_1}{a}\right)e^{-a(t-2T)}\right]u(t-2T) \tag{10.15.20}$$

Applying this result to the four cases, we find:

$$\begin{aligned}
V_d(t) &= V_0\,u(t)+V_0\big[\Gamma_R+(1-\Gamma_R)e^{-a(t-2T)}\big]u(t-2T) &&\text{(series } R\text{–}L)\\
V_d(t) &= V_0\,u(t)+V_0\big[-1+(1+\Gamma_R)e^{-a(t-2T)}\big]u(t-2T) &&\text{(parallel } R\text{–}L)\\
V_d(t) &= V_0\,u(t)+V_0\big[1-(1-\Gamma_R)e^{-a(t-2T)}\big]u(t-2T) &&\text{(series } R\text{–}C)\\
V_d(t) &= V_0\,u(t)+V_0\big[\Gamma_R-(1+\Gamma_R)e^{-a(t-2T)}\big]u(t-2T) &&\text{(parallel } R\text{–}C)
\end{aligned} \tag{10.15.21}$$

In a similar fashion, we determine the load voltage:

$$\begin{aligned}
V_L(t) &= V_0\big[(1+\Gamma_R)+(1-\Gamma_R)e^{-a(t-T)}\big]u(t-T) &&\text{(series } R\text{–}L)\\
V_L(t) &= V_0(1+\Gamma_R)e^{-a(t-T)}\,u(t-T) &&\text{(parallel } R\text{–}L)\\
V_L(t) &= V_0\big[2-(1-\Gamma_R)e^{-a(t-T)}\big]u(t-T) &&\text{(series } R\text{–}C)\\
V_L(t) &= V_0(1+\Gamma_R)\big[1-e^{-a(t-T)}\big]u(t-T) &&\text{(parallel } R\text{–}C)
\end{aligned} \tag{10.15.22}$$

Example 10.15.2: We take $V_0=1$, $Z_0=50$, $R=150\ \Omega$, and, as before, $d=1$, $c=1$, $T=1$. We find $\Gamma_R=0.5$. Fig. 10.15.4 shows the voltages $V_d(t)$ and $V_L(t)$ in the four cases.

In all cases, we adjusted L and C such that $a=1$. This gives $L=200$ and $C=1/200$, and $L=37.5$ and $C=1/37.5$, for the series and parallel cases.

Asymptotically, the series R–L and the parallel R–C cases look like a voltage divider $V_d=V_L=V_GR/(R+Z_0)=1.5$, the parallel R–L case looks like a short-circuited load $V_d=V_L=0$, and the series R–C looks like and open circuit so that $V_d=V_L=V_G=2$.

Using the expressions for $V(t,z)$ of Problem 10.40, the MATLAB file `RLCmovie.m` makes a movie of the step input as it propagates to and gets reflected from the reactive load. □

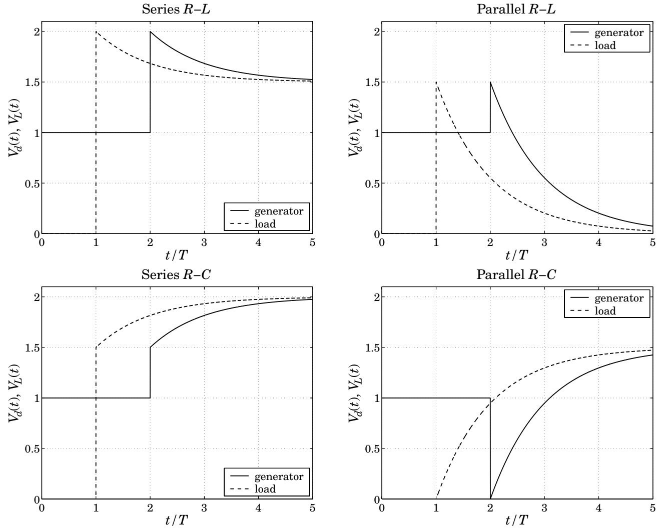


Fig. 10.15.4 Transient response of reactive terminations.

10.16 Problems

10.1 Design a two-wire line made of two AWG 20-gauge (diameter 0.812 mm) copper wires that has a 300-ohm impedance. Calculate its capacitance per unit length.

10.2 For the two-wire line shown in Fig. 10.5.1, show that the tangential component of the electric field vanishes on both cylindrical conductor surfaces. Show that the surface charge and current densities on the positively charged conductor are given in terms of the azimuthal angle ϕ as follows:

$$\rho_s(\phi)=\frac{Q'}{2\pi a}\,\frac{k^2-1}{k^2-2k\cos\phi+1}\,,\qquad J_{sz}(\phi)=\frac{I}{2\pi a}\,\frac{k^2-1}{k^2-2k\cos\phi+1}$$

Show and interpret the following:

$$\int_0^{2\pi} \rho_s(\phi)\, a\, d\phi = Q', \qquad \int_0^{2\pi} J_{sz}(\phi)\, a\, d\phi = I$$

10.3 For the two-wire line of the previous problem, show that the power loss per unit length due to ohmic conductor losses is given by:

$$P'_{\text{loss}} = R_s \int_0^{2\pi} |J_{sz}(\phi)|^2 a\, d\phi = \frac{R_s |I|^2}{2\pi a} \frac{k^2+1}{k^2-1}$$

From this result, derive Eq. (10.5.13) for R' and α_c.

10.4 A polyethylene-filled RG-59 coaxial cable has impedance of 75 ohm and velocity factor of 2/3. If the radius of the inner conductor is 0.322 mm, determine the radius of the outer conductor in mm. Determine the capacitance and inductance per unit length. Assuming copper conductors and a loss tangent of 7×10^{-4} for the polyethylene dielectric, calculate the attenuation of the cable in dB/100-ft at 50 MHz and at 1 GHz. Finally, calculate the cutoff frequency of higher propagating modes.

10.5 *Computer Experiment: Coaxial Cable Attenuation.* Consider the attenuation data of an RG-8/U cable given in Example 10.4.3.

a. Reproduce the graph of that Example. Show that with the assumed characteristics of the cable, the total attenuation may be written as a function of frequency in the form, where α is in dB per 100 ft and f is in GHz:

$$\alpha(f) = 4.3412 f^{1/2} + 2.9131 f$$

b. Carry out a least-squares fit of the attenuation data given in the table of that Example by fitting them to a function of the form $\alpha(f) = A f^{1/2} + B f$, and determine the fitted coefficients A, B. This requires that you find A, B by minimizing the weighted performance index:

$$\mathcal{J} = \sum_i w_i \big(\alpha_i - A f_i^{1/2} - B f_i\big)^2 = \min$$

where you may take the weights $w_i = 1$. Show that the minimization problem gives rise to a 2×2 linear system of equations in the unknowns A, B, and solve this system with MATLAB.

Plot the resulting function of $\alpha(f)$ on the same graph as that of part (a). How do the fitted coefficients compare with those of part (a)?

Given the fitted coefficients A, B, extract from them the estimated values of the loss tangent $\tan\delta$ and the refractive index n of the dielectric filling (assuming the cable radii a, b and conductivity σ are as given.)

c. Because it appears that the 5-GHz data point is not as accurate as the others, redo part (b) by assigning only 1/2 weight to that point in the least-squares fit. Finally, redo part (b) by assigning zero weight to that point (i.e., not using it in the fit.)

10.6 *Computer Experiment—Optimum Coaxial Cables.* Plot the three quantities E_a, P_T, and α_c given in Eq. (10.4.10) versus the ratio b/a over the range $1.5 \le b/a \le 4$. Indicate on the graphs the positions of the optimum ratios that correspond to the minima of E_a and α_c, and the maximum of P_T.

Moreover, write a MATLAB function that solves iteratively (for example, using Newton's method) the equation for minimizing α_c, that is, $\ln x = 1 + 1/x$.

10.7 Let $Z_l = R_l + jX_l$ be the wave impedance on a lossless line at a distance l from a purely resistive load Z_L. Derive explicit expressions for R_l and X_l in terms of Z_L and the characteristic impedance Z_0 of the line for the distances $l = n\lambda/8$, where $n = 1, 2, 3, 4, 5, 6, 7, 8$. Discuss the signs of X_l (inductive or capacitive) for the two cases $Z_L > Z_0$ and $Z_L < Z_0$. What happens to the above expressions when $Z_L = Z_0$?

10.8 A dipole antenna operating in the 30-meter band is connected to a transmitter by a 20-meter long lossless coaxial cable having velocity factor of 0.66 and characteristic impedance of 50 ohm. The wave impedance at the transmitter end of the cable is measured and found to be $39.9 + 34.2j$ ohm. Determine the input impedance of the antenna.

10.9 It is desired to measure the characteristic impedance Z_0 and propagation constant $\gamma = \alpha + j\beta$ of a lossy line. To this end, a length l of the line is short-circuited and its input impedance Z_{sc} is measured. Then, the segment is open-circuited and its input impedance Z_{oc} is measured. Explain how to extract the two unknown quantities Z_0 and γ from Z_{sc} and Z_{oc}.

10.10 The wave impedances of a 100-meter long short- and open-circuited segment of a lossy transmission line were measured to be $Z_{\text{sc}} = 68.45 + 128.13j$ ohm and $Z_{\text{oc}} = 4.99 - 16.65j$ ohm at 10 MHz. Using the results of the previous problem, determine the characteristic impedance of the line Z_0, the attenuation constant α in dB/100-m, and the velocity factor of the cable noting that the cable length is at least two wavelengths long.

10.11 For a lossless line, show the inequality:

$$\frac{1-|\Gamma_L|}{1+|\Gamma_L|} \le \left| \frac{1+\Gamma_L e^{-2j\beta l}}{1-\Gamma_L e^{-2j\beta l}} \right| \le \frac{1+|\Gamma_L|}{1-|\Gamma_L|}$$

where Γ_L is the load reflection coefficient. Then, show that the magnitude of the wave impedance Z_l along the line varies between the limits:

$$Z_{\min} \le |Z_l| \le Z_{\max}, \qquad Z_{\min} = \frac{1}{S} Z_0, \quad Z_{\max} = S\, Z_0$$

where Z_0 is the characteristic impedance of the line and S, the voltage SWR.

10.12 For a lossless line, show that the current I_l at a distance l from a load varies between the limits:

$$I_{\min} \le |I_l| \le I_{\max}, \qquad \text{where} \quad I_{\min} = \frac{1}{Z_0} V_{\min}, \quad I_{\max} = \frac{1}{Z_0} V_{\max}$$

where $V_{\min}$ and $V_{\max}$ are the minimum and maximum voltage along the line. Then, show that the minimum and maximum wave impedances of the previous problem can be written in the alternative forms:

$$Z_{\max} = \frac{V_{\max}}{I_{\min}}, \qquad Z_{\min} = \frac{V_{\min}}{I_{\max}}$$

Recall from Sec. 10.13 that $Z_{\max}$, $Z_{\min}$ correspond to the distances $l_{\max}$ and $l_{\min}$. However, show that $I_{\min}$ and $I_{\max}$ correspond to $l_{\max}$ and $l_{\min}$, respectively.

10.13 If 500 W of power are delivered to a load by a 50-ohm lossless line and the SWR on the line is 5, determine the maximum voltage $V_{\max}$ along the line. Determine also the quantities $V_{\min}$, $I_{\max}$, $I_{\min}$, $Z_{\max}$, and $Z_{\min}$.

10.14 A transmitter is connected to an antenna by an 80-ft length of coaxial cable of characteristic impedance of 50 ohm and matched-line loss of 0.6 dB/100-ft. The antenna impedance is $30 + 40j$ ohm. The transmitter delivers 1 kW of power into the line. Calculate the amount of power delivered to the load and the power lost in the line. Calculate the SWR at the antenna and transmitter ends of the line.

10.15 Let S_L and S_d be the SWRs at the load and at distance d from the load on a lossy and mismatched line. Let $a = e^{2\alpha d}$ be the matched-line loss for the length-d segment. Show that the SWRs are related by:

$$S_d = S_L - \frac{(a-1)(S_L^2-1)}{a(S_L+1)-(S_L-1)} \quad \text{and} \quad S_L = S_d + \frac{(a-1)(S_d^2-1)}{(S_d+1)-a(S_d-1)}$$

Show that $1 \le S_d \le S_L$. When are the equalities valid? Show also that $S_d \to 1$ as $d \to \infty$.

10.16 A 100-Ω lossless transmission line is terminated at an unknown load impedance. The line is operated at a frequency corresponding to a wavelength $\lambda = 40$ cm. The standing wave ratio along this line is measured to be $S = 3$. The distance from the load where there is a voltage minimum is measured to be 5 cm. Based on these two measurements, determine the unknown load impedance.

10.17 The wavelength on a 50 Ω transmission line is 80 cm. Determine the load impedance if the SWR on the line is 3 and the location of the first voltage minimum is 10 cm from the load. At what other distances from the load would one measure a voltage minimum? A voltage maximum?

10.18 A 75-ohm line is connected to an unknown load. Voltage measurements along the line reveal that the maximum and minimum voltage values are 6 V and 2 V. It is observed that a voltage maximum occurs at the distance from the load:

$$l = 0.5\lambda - \frac{\lambda}{4\pi}\,\text{atan}(0.75) = 0.44879\lambda$$

Determine the reflection coefficient Γ_L (in cartesian form) and the load impedance Z_L.

10.19 A load is connected to a generator by a 30-ft long 75-ohm RG-59/U coaxial cable. The SWR is measured at the load and the generator and is found to be equal to 3 and 2, respectively. Determine the attenuation of the cable in dB/ft. Assuming the load is resistive, what are all possible values of the load impedance in ohm?

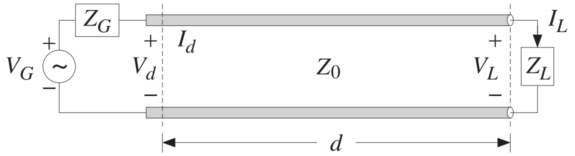
10.20 A lossless 50-ohm line with velocity factor of 0.8 is connected to an unknown load. The operating frequency is 1 GHz. Voltage measurements along the line reveal that the maximum and minimum voltage values are 6 V and 2 V. It is observed that a voltage minimum occurs at a distance of 3 cm from the load. Determine the load reflection coefficient Γ_L and the load impedance Z_L.

10.21 The SWR on a lossy line is measured to be equal to 3 at a distance of 5 meters from the load, and equal to 4 at a distance of 1 meter from the load.

a. Determine the attenuation constant of the line in dB/m.

b. Assuming that the load is purely resistive, determine the two possible values of the load impedance.

10.22 A lossless 50-ohm transmission line of length $d = 17$ m is connected to an unknown load Z_L and to a generator $V_G = 10$ volts having an unknown internal impedance Z_G, as shown below. The wavelength on the line is $\lambda = 8$ m. The current and voltage on the line at the generator end are measured and found to be $I_d = 40$ mA and $V_d = 6$ volts.



a. Determine the wave impedance Z_d at the generator end, as well as the generator's internal impedance Z_G.

b. Determine the load impedance Z_L.

c. What percentage of the total power produced by the generator is absorbed by the load?

10.23 The wavelength on a 50-ohm transmission line is 8 meters. Determine the load impedance if the SWR on the line is 3 and the location of the first voltage maximum is 1 meter from the load. At what other distances from the load would one measure a voltage minimum? A voltage maximum?

10.24 A 10-volt generator with a 25-ohm internal impedance is connected to a 100-ohm load via a 6-meter long 50-ohm transmission line. The wavelength on the line is 8 meters. Carry out the following calculations in the stated order:

a. Calculate the wave impedance Z_d at the generator end of the line. Then, using an equivalent voltage divider circuit, calculate the voltage and current V_d, I_d. Then, calculate the forward and backward voltages V_{d+}, V_{d-} from the knowledge of V_d, I_d.

b. Propagate V_{d+}, V_{d-} to the load end of the line to determine the values of the forward and backward voltages V_{L+}, V_{L-} at the load end. Then, calculate the corresponding voltage and current V_L, I_L from the knowledge of V_{L+}, V_{L-}.

c. Assuming that the real-valued form of the generator voltage is

$$V_G = 10\cos(\omega t)$$

determine the real-valued forms of the quantities V_d, V_L expressed in the sinusoidal form $A\cos(\omega t + \theta)$.

10.25 A lossless 50-ohm transmission line is connected to an unknown load impedance Z_L. Voltage measurements along the line reveal that the maximum and minimum voltage values are $(\sqrt{2}+1)$ volts and $(\sqrt{2}-1)$ volts. Moreover, a distance at which a voltage maximum is observed has been found to be $l_{\max} = 15\lambda/16$.

a. Determine the load reflection coefficient Γ_L and the impedance Z_L.

b. Determine a distance (in units of λ) at which a voltage minimum will be observed.

10.26 A 50-ohm transmission line is terminated at a load impedance:

$$Z_L = 75 + j25\ \Omega$$

a. What percentage of the incident power is reflected back into the line?

b. In order to make the load reflectionless, a short-circuited 50-ohm stub of length d is inserted in parallel at a distance l from the load. What are the smallest values of the lengths d and l in units of the wavelength λ that will make the load reflectionless? Show all work.

10.27 A load is connected to a generator by a 20-meter long 50-ohm coaxial cable. The SWR is measured at the load and the generator and is found to be equal to 3 and 2, respectively.

a. Determine the attenuation of the cable in dB/m.

b. Assuming that the load is resistive, what are all possible values of the load impedance in ohm? [*Hint:* the load impedance can be greater or less than the cable impedance.]

10.28 A 50-ohm lossless transmission line with velocity factor of 0.8 and operating at a frequency of 15 MHz is connected to an unknown load impedance. The voltage SWR is measured to be $S = 3 + 2\sqrt{2}$. A voltage maximum is found at a distance of 1 m from the load.

a. Determine the unknown load impedance Z_L.

b. Suppose that the line is lossy and that it is connected to the load found in part (a). Suppose that the SWR at a distance of 10 m from the load is measured to be $S = 3$. What is the attenuation of the line in dB/m?

10.29 A lossless 50-ohm transmission line is connected to an unknown load impedance. Voltage measurements along the line reveal that the maximum and minimum voltage values are 6 V and 2 V. Moreover, the closest distance to the load at which a voltage minimum is observed has been found to be such that: $e^{2j\beta l_{\min}} = 0.6-0.6j$. Determine the load reflection coefficient Γ_L and the impedance Z_L.

10.30 A resonant dipole antenna operating in the 30-meter band is connected to a transmitter by a 30-meter long lossless coaxial cable having velocity factor of 0.8 and characteristic impedance of 50 ohm. The wave impedance at the transmitter end of the cable is measured to be 40 ohm. Determine the input impedance of the antenna.

10.31 The next four problems are based on Ref. [945]. A lossless transmission line with real characteristic impedance Z_0 is connected to a series RLC circuit.

a. Show that the corresponding load impedance may be written as a function of frequency in the form (with f, f_0 in Hz):

$$Z_L = R + jRQ\left(\frac{f}{f_0} - \frac{f_0}{f}\right)$$

where f_0 and Q are the frequency and Q-factor at resonance. Such a load impedance provides a simplified model for the input impedance of a resonant dipole antenna.

Show that the corresponding SWR S_L satisfies $S_L \geq S_0$ for all f, where S_0 is the SWR at resonance, that is, corresponding to $Z_L = R$.

b. The SWR bandwidth is defined by $\Delta f = f_2 - f_1$, where f_1, f_2 are the left and right bandedge frequencies at which the SWR S_L reaches a certain level, say $S_L = S_B$, such that $S_B > S_0$. Often the choice $S_B = 2$ is made. Assuming that $Z_0 \geq R$, show that the bandedge frequencies satisfy the conditions:

$$f_1 f_2 = f_0^2\,, \qquad f_1^2 + f_2^2 = 2f_0^2 + f_0^2\,\frac{(S_0+1)^2\Gamma_B^2 - (S_0-1)^2}{Q^2(1-\Gamma_B^2)}\,, \quad \text{where} \quad \Gamma_B = \frac{S_B - 1}{S_B + 1}$$

c. Show that the normalized bandwidth is given by:

$$Q\,\frac{\Delta f}{f_0} = \sqrt{(S_B - S_0)\,(S_0 - S_B^{-1})} = \sqrt{\frac{4\,(\Gamma_B^2 - \Gamma_0^2)}{(1-\Gamma_0)^2\,(1-\Gamma_B^2)}}\,, \quad \text{with} \quad \Gamma_0 = \frac{S_0 - 1}{S_0 + 1}$$

Show that the left and right bandedge frequencies are given by:

$$f_1 = \sqrt{f_0^2 + \frac{(\Delta f)^2}{4}} - \frac{\Delta f}{2}\,, \qquad f_2 = \sqrt{f_0^2 + \frac{(\Delta f)^2}{4}} + \frac{\Delta f}{2}$$

d. Show that the maximum bandwidth is realized for a *mismatched* load that has the following optimum SWR at resonance:

$$S_0 = \frac{S_B + S_B^{-1}}{2}\,, \quad \Gamma_0 = \Gamma_B^2 \quad \Rightarrow \quad Q\,\frac{\Delta f_{\max}}{f_0} = \frac{S_B^2 - 1}{2S_B} = \frac{2\Gamma_B}{1 - \Gamma_B^2}$$

For example, if $S_B = 2$, we have $\Gamma_B = 1/3$, $S_0 = 1.25$, and $\Delta f/f_0 = 0.75/Q$, whereas for a matched load we have $S_0 = 1$ and $\Delta f/f_0 = 0.50/Q$.

10.32 We assume now that the transmission line of the previous problem is lossy and that the RLC load is connected to a generator by a length-d segment of the line. Let $a = e^{2\alpha d}$ be the matched-line loss. For such lossy line, we may define the bandwidth in terms of the SWR S_d at the generator end.

Show that the normalized bandwidth is given by the same expression as in the previous problem, but with the replacement $\Gamma_B \to \Gamma_{LB}$, where $\Gamma_{LB} \equiv a\Gamma_B$:

$$Q\,\frac{\Delta f}{f_0} = \sqrt{(S_{LB} - S_0)\,(S_0 - S_{LB}^{-1})} = \sqrt{\frac{4\,(\Gamma_{LB}^2 - \Gamma_0^2)}{(1-\Gamma_0)^2\,(1-\Gamma_{LB}^2)}}\,, \quad \text{where} \quad S_{LB} = \frac{1 + \Gamma_{LB}}{1 - \Gamma_{LB}}$$

Show that Γ_{LB}, S_{LB} are the quantities Γ_B, S_B referred to the load end of the line. Show that the meaningful range of the bandwidth formula is $1 \leq S_0 \leq S_{LB}$ in the lossy case, and $1 \leq S_o \leq S_B$ for the lossless case. Show that for the same S_0 the bandwidth for the lossy case is always greater than the bandwidth of the lossless case.

Show that this definition of bandwidth makes sense as long as the matched line loss satisfies $a\Gamma_B < 1$. Show that the bandwidth vanishes at the S_0 that has $\Gamma_0 = a\Gamma_B$. Show that the maximum bandwidth is realized for the optimum S_0:

$$S_0 = \frac{S_{LB} + S_{LB}^{-1}}{2}\,, \quad \Gamma_0 = \Gamma_{LB}^2 \quad \Rightarrow \quad Q\,\frac{\Delta f_{\max}}{f_0} = \frac{S_{LB}^2 - 1}{2S_{LB}} = \frac{2\Gamma_{LB}}{1 - \Gamma_{LB}^2} = \frac{2a\Gamma_B}{1 - a^2\Gamma_B^2}$$

Show that the optimum S_0 is given at the load and generator ends of the line by:

$$S_0 = \frac{1 + a^2\Gamma_B^2}{1 - a^2\Gamma_B^2}\,, \quad S_{d0} = \frac{1 + a\Gamma_B^2}{1 - a\Gamma_B^2}$$

10.33 Assume now that $Z_0 \leq R$ in the previous problem. Show that the normalized bandwidth is given by:

$$Q\,\frac{\Delta f}{f_0} = \sqrt{(S_{LB} - S_0^{-1})\,(S_0^{-1} - S_{LB}^{-1})} = \sqrt{\frac{4\,(\Gamma_{LB}^2 - \Gamma_0^2)}{(1+\Gamma_0)^2\,(1-\Gamma_{LB}^2)}}$$

Show that the maximum always occurs at $S_0 = 1$. Show that the conditions $a\Gamma_B < 1$ and $0 \leq S_0 \leq S_{LB}$ are still required.

Show that, for the same S_0, the bandwidth of the case $Z_0 \leq R$ is always smaller than that of the case $Z_0 \geq R$.

10.34 *Computer Experiment—Antenna Bandwidth.* An 80-meter dipole antenna is resonant at $f_0 = 3.75$ MHz. Its input impedance is modeled as a series RLC circuit as in Problem 10.31. Its Q-factor is $Q = 13$ and its resistance R at resonance will be varied to achieve various values of the SWR S_0. The antenna is connected to a transmitter with a length of 75-ohm coaxial cable with matched-line loss of $a = e^{2\alpha d}$.

a. For a lossless line ($a = 0$ dB), plot the normalized bandwidths $Q(\Delta f)/f_0$ versus the SWR at the antenna at resonance S_0. Do two such plots corresponding to SWR bandwidth levels of $S_B = 2$ and $S_B = 1.75$. On the same graphs, add the normalized bandwidth plots for the case of a lossy line with $a = 2$ dB. Identify on each graph the optimum bandwidth points and the maximum range of S_0 (for convenience, use the same vertical and horizontal scales in all graphs.)

b. Assume now that $S_0 = 1.25$. What are the two possible values of R? For these two cases and assuming a lossy line with $a = 2$ dB, plot the SWR at the antenna end of the line versus frequency in the interval $3.5 \le f \le 4$ MHz. Then, plot the SWRs at the transmitter end of the line. Using common scales on all four graphs, add on each graph the left and right bandedge frequencies corresponding to the two SWR levels of $S_B = 2$ and $S_B = 1.75$. Note the wider bandwidth in the lossy case and for the case having $Z_0 \ge R$.

10.35 For the special case of a *matched* generator having $Z_L = Z_0$, or, $\Gamma_G = 0$, show that Eq. (10.15.15) reduces to:

$$V_d(t) = V(t) + \Gamma_L V(t-2T) \quad \text{and} \quad V_L(t) = (1+\Gamma_L)V(t-T)$$

10.36 A terminated transmission line may be thought of as a *sampled-data linear system.* Show that Eq. (10.15.15) can be written in the convolutional form:

$$V_d(t) = \int_{-\infty}^{\infty} h_d(t')V(t-t')\,dt', \quad V_L(t) = \int_{-\infty}^{\infty} h_L(t')V(t-t')\,dt'$$

so that $V(t)$ may be considered to be the input and $V_d(t)$ and $V_L(t)$, the outputs. Show that the corresponding impulse responses have the sampled-data forms:

$$h_d(t) = \delta(t) + \left(1 + \frac{1}{\Gamma_G}\right)\sum_{m=1}^{\infty}(\Gamma_G\Gamma_L)^m\,\delta(t-2mT)$$

$$h_L(t) = (1+\Gamma_L)\sum_{m=0}^{\infty}(\Gamma_G\Gamma_L)^m\,\delta\big(t-(2m+1)T\big)$$

What are the corresponding frequency responses? Show that the effective *time constant* of the system may be defined as:

$$\tau = 2T\,\frac{\ln\epsilon}{\ln|\Gamma_G\Gamma_L|}$$

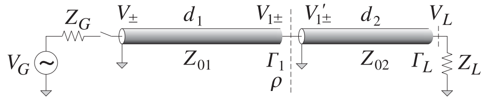
where ϵ is a small number, such as $\epsilon = 10^{-2}$. Provide an interpretation of τ.

10.37 *Computer Experiment—Rise Time and Propagation Effects.* In digital systems where pulses are transmitted along various interconnects, a *rule of thumb* is used according to which if the rise time-constant of a pulse is $t_r \le 2.5T$, where $T = d/c$ is the propagation delay along the interconnect, then propagation effects must be taken into account. If $t_r > 5T$, then a lumped circuit approach may be used.

Consider the transmission line of Example 10.15.1. Using the MATLAB function `upulse.m`, generate four trapezoidal pulses of duration $t_d = 20T$ and rise times $t_r = 0,\ 2.5T,\ 5T,\ 10T$. You may take the fall-times to be equal to the rise-times.

For each pulse, calculate and plot the line voltages $V_d(t), V_L(t)$ at the generator and load ends for the time period $0 \le t \le 80T$. Superimpose on these graphs the initial trapezoidal waveform that is launched along the line. Discuss the above rule of thumb in the light of your results.

10.38 Two coaxial transmission lines of lengths d_1, d_2, impedances Z_{01}, Z_{02}, and propagation speeds c_1, c_2 are connected in cascade as shown below. Define the one-way travel times and z-transform variables by $T_1 = d_1/c_1$, $T_2 = d_2/c_2$, $\zeta_1 = e^{j\omega T_1}$, and $\zeta_2 = e^{j\omega T_2}$.



Show that the reflection response at the left of the junction is given by:

$$\Gamma_1 = \frac{\rho + \Gamma_L\zeta_2^{-2}}{1+\rho\Gamma_L\zeta_2^{-2}} = \rho + \frac{\Gamma_L(1-\rho^2)\zeta_2^{-2}}{1+\rho\Gamma_L\zeta_2^{-2}}$$

where $\rho = (Z_{02}-Z_{01})/(Z_{02}+Z_{01})$ and Γ_L is the load reflection coefficient. Show that the forward and backward voltages at the generator end and to the right of the junction are:

$$V_+ = \frac{V}{1-\Gamma_G\Gamma_1\zeta_1^{-2}}, \quad V_- = \Gamma_1\zeta_1^{-2}V_+, \quad \text{where} \quad V = \frac{V_G Z_{01}}{Z_G + Z_{01}}$$

$$V'_{1+} = \frac{(1+\rho)\zeta_1^{-1}}{1+\rho\Gamma_L\zeta_2^{-2}}V_+, \quad V'_{1-} = \frac{(1+\rho)\Gamma_L\zeta_1^{-1}\zeta_2^{-2}}{1+\rho\Gamma_L\zeta_2^{-2}}V_+$$

Assume a matched generator, that is, having $Z_G = Z_{01}$, or, $\Gamma_G = 0$, and a purely resistive load. Show that the time-domain forward and backward transient voltages are given by:

$$\begin{aligned}
V_+(t) &= V(t) = \frac{1}{2}V_G(t)\\
V_-(t) &= \rho V(t-2T_1) + \Gamma_L(1-\rho^2)\sum_{m=0}^{\infty}(-\rho\Gamma_L)^m V(t-2mT_2-2T_2-2T_1)\\
V'_+(t) &= (1+\rho)\sum_{m=0}^{\infty}(-\rho\Gamma_L)^m V(t-2mT_2-T_1)\\
V'_-(t) &= \Gamma_L(1+\rho)\sum_{m=0}^{\infty}(-\rho\Gamma_L)^m V(t-2mT_2-2T_2-T_1)
\end{aligned}$$

Show that the line voltage $V(t,z)$ is given in terms of the above quantities by:

$$V(t,z) = \begin{cases} V_+(t-z/c_1) + V_-(t+z/c_1), & \text{for} \quad 0 \le z \le d_1\\ V'_{1+}\big(t-(z-d_1)/c_2\big) + V'_{1-}\big(t+(z-d_1)/c_2\big), & \text{for} \quad d_1 \le z \le d_1+d_2 \end{cases}$$

10.39 *Computer Experiment—Transient Response of Cascaded Lines.* For the previous problem, assume the numerical values $d_1 = 8$, $d_2 = 2$, $c_1 = c_2 = 1$, $Z_{01} = 50$, $Z_{02} = 200$, $Z_G = 50$, and $Z_L = 600\ \Omega$.

Plot the line voltage $V_d(t) = V_+(t) + V_-(t)$ at the generator end for $0 \le t \le 5T_1$, in the two cases of (a) a step input $V_G(t) = 3.25\,u(t)$, and (b) a pulse input of width $\tau = T_1/20$ defined by $V_G(t) = 3.25\big[u(t) - u(t-\tau)\big]$. You may use the MATLAB functions `ustep.m` and `upulse.m`.

For case (a), explain also the initial and final voltage levels. In both cases, explain the reasons for the time variations of $V_d(t)$.

The MATLAB file `pulse2movie.m` generates a movie of the pulse or step signal $V(t,z)$ as it propagates through this structure.

10.40 Equations (10.15.21) and (10.15.22) represent the line voltages at the generator and load ends of a line terminated by a reactive load. Using inverse Laplace transforms, show that the line voltage at any point z along such a line is given by:

$$V(t,z)=V_0\,u(t-z/c)+V_0\big[\Gamma_R+(1-\Gamma_R)e^{-a(t+z/c-2T)}\big]u(t+z/c-2T)\qquad\text{(series }R\text{-}L)$$

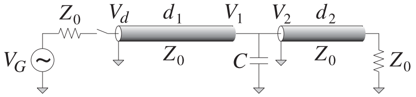
$$V(t,z)=V_0\,u(t-z/c)+V_0\big[-1+(1+\Gamma_R)e^{-a(t+z/c-2T)}\big]u(t+z/c-2T)\qquad\text{(parallel }R\text{-}L)$$

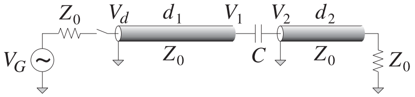
$$V(t,z)=V_0\,u(t-z/c)+V_0\big[1-(1-\Gamma_R)e^{-a(t+z/c-2T)}\big]u(t+z/c-2T)\qquad\text{(series }R\text{-}C)$$

$$V(t,z)=V_0\,u(t-z/c)+V_0\big[\Gamma_R-(1+\Gamma_R)e^{-a(t+z/c-2T)}\big]u(t+z/c-2T)\qquad\text{(parallel }R\text{-}C)$$

The MATLAB file `RLCmovie.m` generates a movie of these waves as they propagate to and get reflected from the reactive load.

10.41 *Time-domain reflectometry* (TDR) is used in a number of applications, such as determining fault locations in buried transmission lines, or probing parts of circuit that would otherwise be inaccessible. As a fault-location example, consider a transmission line of impedance Z_0 matched at both the generator and load ends, having a fault at a distance d_1 from the source, or distance d_2 from the load, as shown below.





The fault is shown as a shunt or series capacitor C. But C can equally well be replaced by an inductor L, or a resistor R. Assuming a unit-step input $V_G(t)=2V_0\,u(t)$, show that the TDR voltage $V_d(t)$ measured at the generator end will be given by:

$$V_d(t)=V_0\,u(t)-V_0\,e^{-a(t-2T_1)}u(t-2T_1)\qquad\text{(shunt }C)$$

$$V_d(t)=V_0\,u(t)-V_0\big[1-e^{-a(t-2T_1)}\big]u(t-2T_1)\qquad\text{(shunt }L)$$

$$V_d(t)=V_0\,u(t)+V_0\big[1-e^{-a(t-2T_1)}\big]u(t-2T_1)\qquad\text{(series }C)$$

$$V_d(t)=V_0\,u(t)+V_0\,e^{-a(t-2T_1)}u(t-2T_1)\qquad\text{(series }L)$$

$$V_d(t)=V_0\,u(t)+V_0\,\Gamma_1\,u(t-2T_1)\qquad\text{(shunt or series }R)$$

where $T_1=d_1/c$ is the one-way travel time to the fault. Show that the corresponding time constant $\tau=1/a$ is in the four cases:

$$\tau=\frac{Z_0C}{2}\,,\quad \tau=2Z_0C\,,\quad \tau=\frac{2L}{Z_0}\,,\quad \tau=\frac{L}{2Z_0}$$

For a resistive fault, show that $\Gamma_1=-Z_0/(2R+Z_0)$, or, $\Gamma_1=R/(2R+Z_0)$, for a shunt or series R. Moreover, show that $\Gamma_1=(Z_1-Z_0)/(Z_1+Z_0)$, where Z_1 is the parallel (in the shunt-R case) or series combination of R with Z_0 and give an intuitive explanation of this fact. For a series C, show that the voltage wave along the two segments is given as follows, and also derive similar expressions for all the other cases:

$$V(t,z)=\begin{cases}V_0\,u(t-z/c)+V_0\big[1-e^{-a(t+z/c-2T_1)}\big]u(t+z/c-2T_1), & \text{for}\quad 0\le z<d_1\\ V_0\,e^{-a(t-z/c)}u(t-z/c), & \text{for}\quad d_1<z\le d_1+d_2\end{cases}$$

Make a plot of $V_d(t)$ for $0\le t\le 5T_1$, assuming $a=1$ for the C and L faults, and $\Gamma_1=\mp1$ corresponding to a shorted shunt or an opened series fault.

The MATLAB file `TDRmovie.m` generates a movie of the step input as it propagates and gets reflected from the fault. The lengths were $d_1=6$, $d_2=4$ (in units such that $c=1$), and the input was $V_0=1$.

11

Coupled Lines

11.1 Coupled Transmission Lines

Coupling between two transmission lines is introduced by their proximity to each other. Coupling effects may be undesirable, such as *crosstalk* in printed circuits, or they may be desirable, as in *directional couplers* where the objective is to transfer power from one line to the other.

In Sections 11.1–11.3, we discuss the equations, and their solutions, describing coupled lines and crosstalk [880–897]. In Sec. 11.4, we discuss directional couplers, as well as fiber Bragg gratings, based on coupled-mode theory [898–919]. Fig. 11.1.1 shows an example of two coupled microstrip lines over a common ground plane, and also shows a generic circuit model for coupled lines.

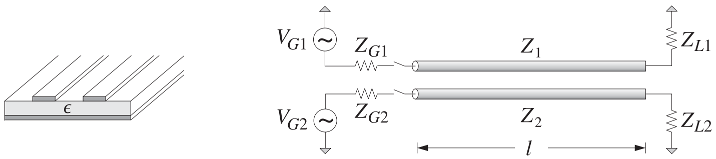


Fig. 11.1.1 Coupled Transmission Lines.

For simplicity, we assume that the lines are lossless. Let L_i, C_i, $i = 1,2$ be the distributed inductances and capacitances *per unit length* when the lines are isolated from each other. The corresponding propagation velocities and characteristic impedances are: $v_i = 1/\sqrt{L_iC_i}$, $Z_i = \sqrt{L_i/C_i}$, $i = 1,2$. The coupling between the lines is modeled by introducing a *mutual* inductance and capacitance per unit length, L_m, C_m. Then, the coupled versions of telegrapher's equations (10.15.1) become:[†]

$$
\begin{aligned}
\frac{\partial V_1}{\partial z} &= -L_1\frac{\partial I_1}{\partial t} - L_m\frac{\partial I_2}{\partial t}\,, &\quad \frac{\partial I_1}{\partial z} &= -C_1\frac{\partial V_1}{\partial t} + C_m\frac{\partial V_2}{\partial t}\\
\frac{\partial V_2}{\partial z} &= -L_2\frac{\partial I_2}{\partial t} - L_m\frac{\partial I_1}{\partial t}\,, &\quad \frac{\partial I_2}{\partial z} &= -C_2\frac{\partial V_2}{\partial t} + C_m\frac{\partial V_1}{\partial t}
\end{aligned}
\qquad (11.1.1)
$$

When $L_m = C_m = 0$, they reduce to the uncoupled equations describing the isolated individual lines. Eqs. (11.1.1) may be written in the 2×2 matrix forms:

$$
\begin{aligned}
\frac{\partial \boldsymbol{V}}{\partial z} &= -\begin{bmatrix} L_1 & L_m \\ L_m & L_2 \end{bmatrix}\frac{\partial \boldsymbol{I}}{\partial t}\\
\frac{\partial \boldsymbol{I}}{\partial z} &= -\begin{bmatrix} C_1 & -C_m \\ -C_m & C_2 \end{bmatrix}\frac{\partial \boldsymbol{V}}{\partial t}
\end{aligned}
\qquad (11.1.2)
$$

where $\boldsymbol{V}, \boldsymbol{I}$ are the column vectors:

$$
\boldsymbol{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad \boldsymbol{I} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \qquad (11.1.3)
$$

For sinusoidal time dependence $e^{j\omega t}$, the system (11.1.2) becomes:

$$
\begin{aligned}
\frac{d\boldsymbol{V}}{dz} &= -j\omega\begin{bmatrix} L_1 & L_m \\ L_m & L_2 \end{bmatrix}\boldsymbol{I}\\
\frac{d\boldsymbol{I}}{dz} &= -j\omega\begin{bmatrix} C_1 & -C_m \\ -C_m & C_2 \end{bmatrix}\boldsymbol{V}
\end{aligned}
\qquad (11.1.4)
$$

It proves convenient to recast these equations in terms of the forward and backward waves that are normalized with respect to the uncoupled impedances Z_1, Z_2:

$$
\begin{aligned}
a_1 &= \frac{V_1 + Z_1I_1}{\sqrt{2Z_1}}\,, & b_1 &= \frac{V_1 - Z_1I_1}{\sqrt{2Z_1}}\\
a_2 &= \frac{V_2 + Z_2I_2}{\sqrt{2Z_2}}\,, & b_2 &= \frac{V_2 - Z_2I_2}{\sqrt{2Z_2}}
\end{aligned}
\quad \Rightarrow \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}
\qquad (11.1.5)
$$

The **a**, **b** waves are similar to the power waves defined in Sec. 13.7. The total average power on the line can be expressed conveniently in terms of these:

$$
\begin{aligned}
P &= \frac{1}{2}\mathrm{Re}[\boldsymbol{V}^\dagger\boldsymbol{I}] = \frac{1}{2}\mathrm{Re}[V_1^*I_1] + \frac{1}{2}\mathrm{Re}[V_2^*I_2] = P_1 + P_2\\
&= \big(|a_1|^2 - |b_1|^2\big) + \big(|a_2|^2 - |b_2|^2\big) = \big(|a_1|^2 + |a_2|^2\big) - \big(|b_1|^2 + |b_2|^2\big)\\
&= \mathbf{a}^\dagger\mathbf{a} - \mathbf{b}^\dagger\mathbf{b}
\end{aligned}
\qquad (11.1.6)
$$

where the dagger operator denotes the conjugate-transpose, for example, $\mathbf{a}^\dagger = [a_1^*, a_2^*]$. Thus, the **a**-waves carry power forward, and the **b**-waves, backward. After some algebra, it can be shown that Eqs. (11.1.4) are equivalent to the system:

$$
\boxed{
\begin{aligned}
\frac{d\mathbf{a}}{dz} &= -jF\,\mathbf{a} + jG\,\mathbf{b}\\
\frac{d\mathbf{b}}{dz} &= -jG\,\mathbf{a} + jF\,\mathbf{b}
\end{aligned}}
\quad \Rightarrow \quad \frac{d}{dz}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = -j\begin{bmatrix} F & -G \\ G & -F \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}
\qquad (11.1.7)
$$

[†] C_1 is related to the capacitance to ground C_{1g} via $C_1 = C_{1g} + C_m$, so that the total charge per unit length on line-1 is $Q_1 = C_1V_1 - C_mV_2 = C_{1g}(V_1 - V_g) + C_m(V_1 - V_2)$, where $V_g = 0$.

with the matrices F, G given by:

$$F = \begin{bmatrix} \beta_1 & \kappa \\ \kappa & \beta_2 \end{bmatrix}, \quad G = \begin{bmatrix} 0 & \chi \\ \chi & 0 \end{bmatrix} \tag{11.1.8}$$

where β_1, β_2 are the uncoupled wavenumbers $\beta_i = \omega/v_i = \omega\sqrt{L_i C_i}$, $i = 1, 2$ and the coupling parameters κ, χ are:

$$\begin{aligned} \kappa &= \frac{1}{2}\omega\left(\frac{L_m}{\sqrt{Z_1 Z_2}} - C_m\sqrt{Z_1 Z_2}\right) = \frac{1}{2}\sqrt{\beta_1\beta_2}\left(\frac{L_m}{\sqrt{L_1 L_2}} - \frac{C_m}{\sqrt{C_1 C_2}}\right) \\ \chi &= \frac{1}{2}\omega\left(\frac{L_m}{\sqrt{Z_1 Z_2}} + C_m\sqrt{Z_1 Z_2}\right) = \frac{1}{2}\sqrt{\beta_1\beta_2}\left(\frac{L_m}{\sqrt{L_1 L_2}} + \frac{C_m}{\sqrt{C_1 C_2}}\right) \end{aligned} \tag{11.1.9}$$

A consequence of the structure of the matrices F, G is that the total power P defined in (11.1.6) is conserved along z. This follows by writing the power in the following form, where I is the 2×2 identity matrix:

$$P = \mathbf{a}^\dagger\mathbf{a} - \mathbf{b}^\dagger\mathbf{b} = [\mathbf{a}^\dagger, \mathbf{b}^\dagger]\begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$$

Using (11.1.7), we find:

$$\frac{dP}{dz} = j[\mathbf{a}^\dagger, \mathbf{b}^\dagger]\left(\begin{bmatrix} F^\dagger & G^\dagger \\ -G^\dagger & -F^\dagger \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} - \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}\begin{bmatrix} F & -G \\ G & -F \end{bmatrix}\right)\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = 0$$

the latter following from the conditions $F^\dagger = F$ and $G^\dagger = G$. Eqs. (11.1.6) and (11.1.7) form the basis of coupled-mode theory.

Next, we specialize to the case of two *identical* lines that have $L_1 = L_2 \equiv L_0$ and $C_1 = C_2 \equiv C_0$, so that $\beta_1 = \beta_2 = \omega\sqrt{L_0 C_0} \equiv \beta$ and $Z_1 = Z_2 = \sqrt{L_0/C_0} \equiv Z_0$, and speed $v_0 = 1/\sqrt{L_0 C_0}$. Then, the $\mathbf{a}, \mathbf{b}$ waves and the matrices F, G take the simpler forms:

$$\mathbf{a} = \frac{\boldsymbol{V} + Z_0\boldsymbol{I}}{\sqrt{2Z_0}}, \quad \mathbf{b} = \frac{\boldsymbol{V} - Z_0\boldsymbol{I}}{\sqrt{2Z_0}} \quad \Rightarrow \quad \mathbf{a} = \frac{\boldsymbol{V} + Z_0\boldsymbol{I}}{2}, \quad \mathbf{b} = \frac{\boldsymbol{V} - Z_0\boldsymbol{I}}{2} \tag{11.1.10}$$

$$F = \begin{bmatrix} \beta & \kappa \\ \kappa & \beta \end{bmatrix}, \quad G = \begin{bmatrix} 0 & \chi \\ \chi & 0 \end{bmatrix} \tag{11.1.11}$$

where, for simplicity, we removed the common scale factor $\sqrt{2Z_0}$ from the denominator of $\mathbf{a}, \mathbf{b}$. The parameters κ, χ are obtained by setting $Z_1 = Z_2 = Z_0$ in (11.1.9):

$$\kappa = \frac{1}{2}\beta\left(\frac{L_m}{L_0} - \frac{C_m}{C_0}\right), \quad \chi = \frac{1}{2}\beta\left(\frac{L_m}{L_0} + \frac{C_m}{C_0}\right), \tag{11.1.12}$$

The matrices F, G commute with each other. In fact, they are both examples of matrices of the form:

$$A = \begin{bmatrix} a_0 & a_1 \\ a_1 & a_0 \end{bmatrix} = a_0 I + a_1 J, \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{11.1.13}$$

where a_0, a_1 are real such that $|a_0| \neq |a_1|$. Such matrices form a *commutative* subgroup of the group of nonsingular 2×2 matrices. Their eigenvalues are $\lambda_\pm = a_0 \pm a_1$ and they can all be diagonalized by a *common* unitary matrix:

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = [\mathbf{e}_+, \mathbf{e}_-], \quad \mathbf{e}_+ = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{e}_- = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{11.1.14}$$

so that we have $QQ^\dagger = Q^\dagger Q = I$ and $A\mathbf{e}_\pm = \lambda_\pm\mathbf{e}_\pm$.

The eigenvectors $\mathbf{e}_\pm$ are referred to as the *even* and *odd* modes. To simplify subsequent expressions, we will denote the eigenvalues of A by $A_\pm = a_0 \pm a_1$ and the diagonalized matrix by $\bar{A}$. Thus,

$$A = Q\bar{A}Q^\dagger, \quad \bar{A} = \begin{bmatrix} A_+ & 0 \\ 0 & A_- \end{bmatrix} = \begin{bmatrix} a_0 + a_1 & 0 \\ 0 & a_0 - a_1 \end{bmatrix} \tag{11.1.15}$$

Such matrices, as well as any matrix-valued function thereof, may be diagonalized simultaneously. Three examples of such functions appear in the solution of Eqs. (11.1.7):

$$\begin{aligned} \mathcal{B} &= \sqrt{(F+G)(F-G)} = Q\sqrt{(\bar{F}+\bar{G})(\bar{F}-\bar{G})}\,Q^\dagger \\ Z &= Z_0\sqrt{(F+G)(F-G)^{-1}} = Z_0 Q\sqrt{(\bar{F}+\bar{G})(\bar{F}-\bar{G})^{-1}}\,Q^\dagger \\ \Gamma &= (Z - Z_0 I)(Z + Z_0 I)^{-1} = Q(\bar{Z} - Z_0 I)(\bar{Z} + Z_0 I)^{-1}Q^\dagger \end{aligned} \tag{11.1.16}$$

Using the property $FG = GF$, and differentiating (11.1.7) one more time, we obtain the decoupled second-order equations, with $\mathcal{B}$ as defined in (11.1.16):

$$\frac{d^2\mathbf{a}}{dz^2} = -\mathcal{B}^2\,\mathbf{a}, \quad \frac{d^2\mathbf{b}}{dz^2} = -\mathcal{B}^2\,\mathbf{b}$$

However, it is better to work with (11.1.7) directly. This system can be decoupled by forming the following linear combinations of the $\mathbf{a}, \mathbf{b}$ waves:

$$\begin{aligned} \boldsymbol{A} &= \mathbf{a} - \Gamma\mathbf{b} \\ \boldsymbol{B} &= \mathbf{b} - \Gamma\mathbf{a} \end{aligned} \quad \Rightarrow \quad \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} I & -\Gamma \\ -\Gamma & I \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{11.1.17}$$

The $\boldsymbol{A}, \boldsymbol{B}$ can be written in terms of $\boldsymbol{V}, \boldsymbol{I}$ and the impedance matrix Z as follows:

$$\begin{aligned} \boldsymbol{A} &= (2D)^{-1}(\boldsymbol{V} + Z\boldsymbol{I}) \\ \boldsymbol{B} &= (2D)^{-1}(\boldsymbol{V} - Z\boldsymbol{I}) \end{aligned} \quad \Rightarrow \quad \begin{aligned} \boldsymbol{V} &= D(\boldsymbol{A} + \boldsymbol{B}) \\ Z\boldsymbol{I} &= D(\boldsymbol{A} - \boldsymbol{B}) \end{aligned} \qquad D = \frac{Z + Z_0 I}{2Z_0} \tag{11.1.18}$$

Using (11.1.17), we find that $\boldsymbol{A}, \boldsymbol{B}$ satisfy the decoupled first-order system:

$$\frac{d}{dz}\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} = -j\begin{bmatrix} \mathcal{B} & 0 \\ 0 & -\mathcal{B} \end{bmatrix}\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} \quad \Rightarrow \quad \frac{d\boldsymbol{A}}{dz} = -j\mathcal{B}\boldsymbol{A}, \quad \frac{d\boldsymbol{B}}{dz} = j\mathcal{B}\boldsymbol{B} \tag{11.1.19}$$

with solutions expressed in terms of the matrix exponentials $e^{\pm j\mathcal{B}z}$:

$$\boldsymbol{A}(z) = e^{-j\mathcal{B}z}\boldsymbol{A}(0), \quad \boldsymbol{B}(z) = e^{j\mathcal{B}z}\boldsymbol{B}(0) \tag{11.1.20}$$

Using (11.1.18), we obtain the solutions for $\boldsymbol{V}, \boldsymbol{I}$:

$$
\begin{aligned}
\boldsymbol{V}(z) &= D\big[e^{-j\mathcal{B}z}\boldsymbol{A}(0) + e^{j\mathcal{B}z}\boldsymbol{B}(0)\big] \\
Z\boldsymbol{I}(z) &= D\big[e^{-j\mathcal{B}z}\boldsymbol{A}(0) - e^{j\mathcal{B}z}\boldsymbol{B}(0)\big]
\end{aligned}
\tag{11.1.21}
$$

To complete the solution, we assume that both lines are terminated at common generator and load impedances, that is, $Z_{G1} = Z_{G2} \equiv Z_G$ and $Z_{L1} = Z_{L2} \equiv Z_L$. The generator voltages V_{G1}, V_{G2} are assumed to be different. We define the generator voltage vector and source and load matrix reflection coefficients:

$$
\boldsymbol{V}_G = \begin{bmatrix} V_{G1} \\ V_{G2} \end{bmatrix}, \qquad
\begin{aligned}
\Gamma_G &= (Z_G I - Z)(Z_G I + Z)^{-1} \\
\Gamma_L &= (Z_L I - Z)(Z_L I + Z)^{-1}
\end{aligned}
\tag{11.1.22}
$$

The terminal conditions for the line are at $z = 0$ and $z = l$:

$$
\boldsymbol{V}_G = \boldsymbol{V}(0) + Z_G\boldsymbol{I}(0)\,, \quad \boldsymbol{V}(l) = Z_L\boldsymbol{I}(l)
\tag{11.1.23}
$$

They may be re-expressed in terms of $\boldsymbol{A}, \boldsymbol{B}$ with the help of (11.1.18):

$$
\boldsymbol{A}(0) - \Gamma_G\boldsymbol{B}(0) = D^{-1}Z(Z + Z_G I)^{-1}\boldsymbol{V}_G\,, \quad \boldsymbol{B}(l) = \Gamma_L\boldsymbol{A}(l)
\tag{11.1.24}
$$

But from (11.1.19), we have:†

$$
e^{j\mathcal{B}l}\boldsymbol{B}(0) = \boldsymbol{B}(l) = \Gamma_L\boldsymbol{A}(l) = \Gamma_L e^{-j\mathcal{B}l}\boldsymbol{A}(0) \quad\Rightarrow\quad \boldsymbol{B}(0) = \Gamma_L e^{-2j\mathcal{B}l}\boldsymbol{A}(0)
\tag{11.1.25}
$$

Inserting this into (11.1.24), we may solve for $\boldsymbol{A}(0)$ in terms of the generator voltage:

$$
\boldsymbol{A}(0) = D^{-1}\big[I - \Gamma_G\Gamma_L e^{-2j\mathcal{B}l}\big]^{-1}Z(Z + Z_G I)^{-1}\boldsymbol{V}_G
\tag{11.1.26}
$$

Using (11.1.26) into (11.1.21), we finally obtain the voltage and current at an arbitrary position z along the lines:

$$
\boxed{
\begin{aligned}
\boldsymbol{V}(z) &= [e^{-j\mathcal{B}z} + \Gamma_L e^{-2j\mathcal{B}l}e^{j\mathcal{B}z}]\,[I - \Gamma_G\Gamma_L e^{-2j\mathcal{B}l}]^{-1}Z(Z + Z_G I)^{-1}\boldsymbol{V}_G \\
\boldsymbol{I}(z) &= [e^{-j\mathcal{B}z} - \Gamma_L e^{-2j\mathcal{B}l}e^{j\mathcal{B}z}]\,[I - \Gamma_G\Gamma_L e^{-2j\mathcal{B}l}]^{-1}(Z + Z_G I)^{-1}\boldsymbol{V}_G
\end{aligned}
}
\tag{11.1.27}
$$

These are the coupled-line generalizations of Eqs. (10.9.7). Resolving $\boldsymbol{V}_G$ and $\boldsymbol{V}(z)$ into their even and odd modes, that is, expressing them as linear combinations of the eigenvectors $\mathbf{e}_\pm$, we have:

$$
\begin{aligned}
&\boldsymbol{V}_G = V_{G+}\mathbf{e}_+ + V_{G-}\mathbf{e}_-\,, \quad \text{where} \quad V_{G\pm} = \frac{V_{G1} \pm V_{G2}}{\sqrt{2}} \\
&\boldsymbol{V}(z) = V_+(z)\mathbf{e}_+ + V_-(z)\mathbf{e}_-\,, \qquad V_\pm(z) = \frac{V_1(z) \pm V_2(z)}{\sqrt{2}}
\end{aligned}
\tag{11.1.28}
$$

In this basis, the matrices in (11.1.27) are diagonal resulting in the equivalent solution:

$$
\begin{aligned}
\boldsymbol{V}(z) = V_+(z)\mathbf{e}_+ + V_-(z)\mathbf{e}_- &= \frac{e^{-j\beta_+z} + \Gamma_{L+}e^{-2j\beta_+l}e^{j\beta_+z}}{1 - \Gamma_{G+}\Gamma_{L+}e^{-2j\beta_+l}}\,\frac{Z_+}{Z_+ + Z_G}\,V_{G+}\mathbf{e}_+ \\
&+ \frac{e^{-j\beta_-z} + \Gamma_{L-}e^{-2j\beta_-l}e^{j\beta_-z}}{1 - \Gamma_{G-}\Gamma_{L-}e^{-2j\beta_-l}}\,\frac{Z_-}{Z_- + Z_G}\,V_{G-}\mathbf{e}_-
\end{aligned}
\tag{11.1.29}
$$

---

† The matrices $D, Z, \Gamma_G, \Gamma_L, \Gamma, \mathcal{B}$ all commute with each other.

where $\beta_\pm$ are the eigenvalues of $\mathcal{B}$, $Z_\pm$ the eigenvalues of Z, and $\Gamma_{G\pm}, \Gamma_{L\pm}$ are:

$$
\Gamma_{G\pm} = \frac{Z_G - Z_\pm}{Z_G + Z_\pm}\,, \quad \Gamma_{L\pm} = \frac{Z_L - Z_\pm}{Z_L + Z_\pm}
\tag{11.1.30}
$$

The voltages $V_1(z), V_2(z)$ are obtained by extracting the top and bottom components of (11.1.29), that is, $V_{1,2}(z) = \big[V_+(z) \pm V_-(z)\big]/\sqrt{2}$:

$$
\begin{aligned}
V_1(z) &= \frac{e^{-j\beta_+z} + \Gamma_{L+}e^{-2j\beta_+l}e^{j\beta_+z}}{1 - \Gamma_{G+}\Gamma_{L+}e^{-2j\beta_+l}}V_+ + \frac{e^{-j\beta_-z} + \Gamma_{L-}e^{-2j\beta_-l}e^{j\beta_-z}}{1 - \Gamma_{G-}\Gamma_{L-}e^{-2j\beta_-l}}V_- \\
V_2(z) &= \frac{e^{-j\beta_+z} + \Gamma_{L+}e^{-2j\beta_+l}e^{j\beta_+z}}{1 - \Gamma_{G+}\Gamma_{L+}e^{-2j\beta_+l}}V_+ - \frac{e^{-j\beta_-z} + \Gamma_{L-}e^{-2j\beta_-l}e^{j\beta_-z}}{1 - \Gamma_{G-}\Gamma_{L-}e^{-2j\beta_-l}}V_-
\end{aligned}
\tag{11.1.31}
$$

where we defined:

$$
V_\pm = \left(\frac{Z_\pm}{Z_\pm + Z_G}\right)\frac{V_{G\pm}}{\sqrt{2}} = \frac{1}{4}(1 - \Gamma_{G\pm})(V_{G1} \pm V_{G2})
\tag{11.1.32}
$$

The parameters $\beta_\pm, Z_\pm$ are obtained using the rules of Eq. (11.1.15). From Eq. (11.1.12), we find the eigenvalues of the matrices $F \pm G$:

$$
\begin{aligned}
(F + G)_\pm &= \beta \pm (\kappa + \chi) = \beta\left(1 \pm \frac{L_m}{L_0}\right) = \omega\frac{1}{Z_0}(L_0 \pm L_m) \\
(F - G)_\pm &= \beta \pm (\kappa - \chi) = \beta\left(1 \mp \frac{C_m}{C_0}\right) = \omega Z_0(C_0 \mp C_m)
\end{aligned}
$$

Then, it follows that:

$$
\begin{aligned}
\beta_+ &= \sqrt{(F + G)_+(F - G)_+} = \omega\sqrt{(L_0 + L_m)(C_0 - C_m)} \\
\beta_- &= \sqrt{(F + G)_-(F - G)_-} = \omega\sqrt{(L_0 - L_m)(C_0 + C_m)}
\end{aligned}
\tag{11.1.33}
$$

$$
\begin{aligned}
Z_+ &= Z_0\sqrt{\frac{(F + G)_+}{(F - G)_+}} = \sqrt{\frac{L_0 + L_m}{C_0 - C_m}} \\
Z_- &= Z_0\sqrt{\frac{(F + G)_-}{(F - G)_-}} = \sqrt{\frac{L_0 - L_m}{C_0 + C_m}}
\end{aligned}
\tag{11.1.34}
$$

Thus, the coupled system acts as two uncoupled lines with wavenumbers and characteristic impedances $\beta_\pm, Z_\pm$, propagation speeds $v_\pm = 1/\sqrt{(L_0 \pm L_m)(C_0 \mp C_m)}$, and propagation delays $T_\pm = l/v_\pm$. The even mode is energized when $V_{G2} = V_{G1}$, or, $V_{G+} \neq 0, V_{G-} = 0$, and the odd mode, when $V_{G2} = -V_{G1}$, or, $V_{G+} = 0, V_{G-} \neq 0$.

When the coupled lines are immersed in a *homogeneous* medium, such as two parallel wires in air over a ground plane, then the propagation speeds must be equal to the speed of light within this medium [890], that is, $v_+ = v_- = 1/\sqrt{\mu\epsilon}$. This requires:

$$
\begin{aligned}
(L_0 + L_m)(C_0 - C_m) &= \mu\epsilon \\
(L_0 - L_m)(C_0 + C_m) &= \mu\epsilon
\end{aligned}
\quad\Rightarrow\quad
\begin{aligned}
L_0 &= \frac{\mu\epsilon C_0}{C_0^2 - C_m^2} \\
L_m &= \frac{\mu\epsilon C_m}{C_0^2 - C_m^2}
\end{aligned}
\tag{11.1.35}
$$

Therefore, $L_m/L_0 = C_m/C_0$, or, equivalently, $\kappa = 0$. On the other hand, in an *inhomogeneous* medium, such as for the case of the microstrip lines shown in Fig. 11.1.1, the propagation speeds may be different, $v_+ \neq v_-$, and hence $T_+ \neq T_-$.

11.2 Crosstalk Between Lines

When only line-1 is energized, that is, $V_{G1} \neq 0, V_{G2} = 0$, the coupling between the lines induces a propagating wave in line-2, referred to as crosstalk, which also has some minor influence back on line-1. The *near-end* and *far-end* crosstalk are the values of $V_2(z)$ at $z = 0$ and $z = l$, respectively. Setting $V_{G2} = 0$ in (11.1.32), we have from (11.1.31):

$$\begin{aligned} V_2(0) &= \frac{1}{2}\frac{(1-\Gamma_{G+})(1+\Gamma_{L+}\zeta_+^{-2})}{1-\Gamma_{G+}\Gamma_{L+}\zeta_+^{-2}}V - \frac{1}{2}\frac{(1-\Gamma_{G-})(1+\Gamma_{L-}\zeta_-^{-1})}{1-\Gamma_{G-}\Gamma_{L-}\zeta^{-2}}V \\ V_2(l) &= \frac{1}{2}\frac{\zeta_+^{-1}(1-\Gamma_{G+})(1+\Gamma_{L+})}{1-\Gamma_{G+}\Gamma_{L+}\zeta_+^{-2}}V - \frac{1}{2}\frac{\zeta_-^{-1}(1-\Gamma_{G-})(1+\Gamma_{L-})}{1-\Gamma_{G-}\Gamma_{L-}\zeta_-^{-2}}V \end{aligned} \tag{11.2.1}$$

where we defined $V = V_{G1}/2$ and introduced the z-transform delay variables $\zeta_\pm = e^{j\omega T_\pm} = e^{j\beta_\pm l}$. Assuming purely resistive termination impedances Z_G, Z_L, we may use Eq. (10.15.15) to obtain the corresponding time-domain responses:

$$\begin{aligned} V_2(0,t) &= \frac{1}{2}(1-\Gamma_{G+})\left[V(t) + \left(1+\frac{1}{\Gamma_{G+}}\right)\sum_{m=1}^{\infty}(\Gamma_{G+}\Gamma_{L+})^m V(t-2mT_+)\right] \\ &\quad - \frac{1}{2}(1-\Gamma_{G-})\left[V(t) + \left(1+\frac{1}{\Gamma_{G-}}\right)\sum_{m=1}^{\infty}(\Gamma_{G-}\Gamma_{L-})^m V(t-2mT_-)\right] \\ V_2(l,t) &= \frac{1}{2}(1-\Gamma_{G+})(1+\Gamma_{L+})\sum_{m=0}^{\infty}(\Gamma_{G+}\Gamma_{L+})^m V(t-2mT_+-T_+) \\ &\quad - \frac{1}{2}(1-\Gamma_{G-})(1+\Gamma_{L-})\sum_{m=0}^{\infty}(\Gamma_{G-}\Gamma_{L-})^m V(t-2mT_--T_-) \end{aligned} \tag{11.2.2}$$

where $V(t) = V_{G1}(t)/2$.[†] Because $Z_\pm \neq Z_0$, there will be multiple reflections even when the lines are matched to Z_0 at both ends. Setting $Z_G = Z_L = Z_0$, gives for the reflection coefficients (11.1.30):

$$\Gamma_{G\pm} = \Gamma_{L\pm} = \frac{Z_0 - Z_\pm}{Z_0 + Z_\pm} = -\Gamma_\pm \tag{11.2.3}$$

In this case, we find for the crosstalk signals:

$$\begin{aligned} V_2(0,t) &= \frac{1}{2}(1+\Gamma_+)\left[V(t) - (1-\Gamma_+)\sum_{m=1}^{\infty}\Gamma_+^{2m-1}V(t-2mT_+)\right] \\ &\quad - \frac{1}{2}(1+\Gamma_-)\left[V(t) - (1-\Gamma_-)\sum_{m=1}^{\infty}\Gamma_-^{2m-1}V(t-2mT_-)\right] \\ V_2(l,t) &= \frac{1}{2}(1-\Gamma_+^2)\sum_{m=0}^{\infty}\Gamma_+^{2m}V(t-2mT_+-T_+) \\ &\quad - \frac{1}{2}(1-\Gamma_-^2)\sum_{m=0}^{\infty}\Gamma_-^{2m}V(t-2mT_--T_-) \end{aligned} \tag{11.2.4}$$

[†] $V(t)$ is the signal that would exist on a matched line-1 in the absence of line-2, $V = Z_0V_{G1}/(Z_0+Z_G) = V_{G1}/2$, provided $Z_G = Z_0$.

Similarly, the near-end and far-end signals on the driven line are found by adding, instead of subtracting, the even- and odd-mode terms:

$$\begin{aligned} V_1(0,t) &= \frac{1}{2}(1+\Gamma_+)\left[V(t) - (1-\Gamma_+)\sum_{m=1}^{\infty}\Gamma_+^{2m-1}V(t-2mT_+)\right] \\ &\quad + \frac{1}{2}(1+\Gamma_-)\left[V(t) - (1-\Gamma_-)\sum_{m=1}^{\infty}\Gamma_-^{2m-1}V(t-2mT_-)\right] \\ V_1(l,t) &= \frac{1}{2}(1-\Gamma_+^2)\sum_{m=0}^{\infty}\Gamma_+^{2m}V(t-2mT_+-T_+) \\ &\quad + \frac{1}{2}(1-\Gamma_-^2)\sum_{m=0}^{\infty}\Gamma_-^{2m}V(t-2mT_--T_-) \end{aligned} \tag{11.2.5}$$

These expressions simplify drastically if we assume weak coupling. It is straightforward to verify that to *first-order* in the parameters $L_m/L_0, C_m/C_0$, or equivalently, to first-order in κ, χ, we have the approximations:

$$\begin{aligned} &\beta_\pm = \beta \pm \Delta\beta = \beta \pm \kappa\,, \quad Z_\pm = Z_0 \pm \Delta Z = Z_0 \pm Z_0\frac{\chi}{\beta}\,, \quad v_\pm = v_0 \mp v_0\frac{\kappa}{\beta} \\ &\Gamma_\pm = 0 \pm \Delta\Gamma = \pm\frac{\chi}{2\beta}\,, \quad T_\pm = T \pm \Delta T = T \pm T\frac{\kappa}{\beta} \end{aligned} \tag{11.2.6}$$

where $T = l/v_0$. Because the $\Gamma_\pm$s are already first-order, the multiple reflection terms in the above summations are a second-order effect, and only the lowest terms will contribute, that is, the term $m = 1$ for the near-end, and $m = 0$ for the far end. Then,

$$\begin{aligned} V_2(0,t) &= \frac{1}{2}(\Gamma_+ - \Gamma_-)V(t) - \frac{1}{2}\big[\Gamma_+V(t-2T_+) - \Gamma_-V(t-2T_-)\big] \\ V_2(l,t) &= \frac{1}{2}\big[V(t-T_+) - V(t-T_-)\big] \end{aligned}$$

Using a Taylor series expansion and (11.2.6), we have to first-order:

$$\begin{aligned} V(t-2T_\pm) &= V(t-2T\mp\Delta T) \simeq V(t-2T) \mp (\Delta T)\dot{V}(t-2T)\,, \quad \dot{V} = \frac{dV}{dt} \\ V(t-T_\pm) &= V(t-T\mp\Delta T) \simeq V(t-T) \mp (\Delta T)\dot{V}(t-T) \end{aligned}$$

Therefore, $\Gamma_\pm V(t-2T_\pm) = \Gamma_\pm\big[V(t-2T) \mp (\Delta T)\dot{V}\big] \simeq \Gamma_\pm V(t-2T)$, where we ignored the second-order terms $\Gamma_\pm(\Delta T)\dot{V}$. It follows that:

$$\begin{aligned} V_2(0,t) &= \frac{1}{2}(\Gamma_+ - \Gamma_-)\big[V(t) - V(t-2T)\big] = (\Delta\Gamma)\big[V(t) - V(t-2T)\big] \\ V_2(l,t) &= \frac{1}{2}\big[V(t-T) - (\Delta T)\dot{V} - V(t-T) - (\Delta T)\dot{V}\big] = -(\Delta T)\frac{dV(t-T)}{dt} \end{aligned}$$

These can be written in the commonly used form:

$$\boxed{\begin{aligned} V_2(0,t) &= K_b\big[V(t) - V(t-2T)\big] \\ V_2(l,t) &= K_f\frac{dV(t-T)}{dt} \end{aligned}} \qquad \text{(near- and far-end crosstalk)} \tag{11.2.7}$$

where K_b, K_f are known as the backward and forward crosstalk coefficients:

$$K_b = \frac{\chi}{2\beta} = \frac{v_0}{4}\left(\frac{L_m}{Z_0} + C_m Z_0\right), \quad K_f = -T\,\frac{\kappa}{\beta} = -\frac{v_0 T}{2}\left(\frac{L_m}{Z_0} - C_m Z_0\right) \tag{11.2.8}$$

where we may replace $l = v_0 T$. The same approximations give for line-1, $V_1(0,t) = V(t)$ and $V_1(l,t) = V(t-T)$. Thus, to first-order, line-2 does not act back to disturb line-1.

Example 11.2.1: Fig. 11.2.1 shows the signals $V_1(0,t)$, $V_1(l,t)$, $V_2(0,t)$, $V_2(l,t)$ for a pair of coupled lines matched at both ends. The uncoupled line impedance was $Z_0 = 50\ \Omega$.

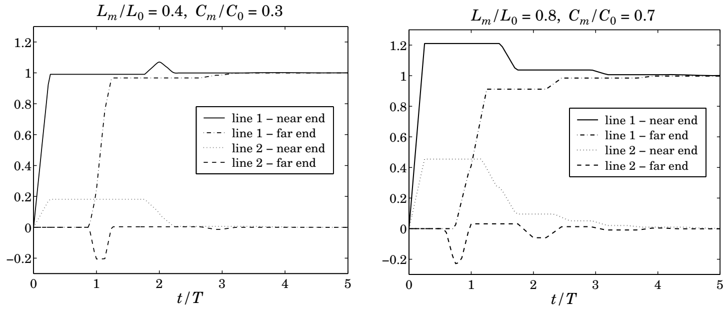


Fig. 11.2.1 Near- and far-end crosstalk signals on lines 1 and 2.

For the left graph, we chose $L_m/L_0 = 0.4$, $C_m/C_0 = 0.3$, which results in the even and odd mode parameters (using the exact formulas):

$$Z_+ = 70.71\ \Omega, \quad Z_- = 33.97\ \Omega, \quad v_+ = 1.01v_0, \quad v_- = 1.13v_0$$
$$\Gamma_+ = 0.17, \quad \Gamma_- = -0.19, \quad T_+ = 0.99T, \quad T_- = 0.88T, \quad K_b = 0.175, \quad K_f = 0.05$$

The right graph corresponds to $L_m/L_0 = 0.8$, $C_m/C_0 = 0.7$, with parameters:

$$Z_+ = 122.47\ \Omega, \quad Z_- = 17.15\ \Omega, \quad v_+ = 1.36v_0, \quad v_- = 1.71v_0$$
$$\Gamma_+ = 0.42, \quad \Gamma_- = -0.49, \quad T_+ = 0.73T, \quad T_- = 0.58T, \quad K_b = 0.375, \quad K_f = 0.05$$

The generator input to line-1 was a rising step with rise-time $t_r = T/4$, that is,

$$V(t) = \frac{1}{2}V_{G1}(t) = \frac{t}{t_r}\big[u(t) - u(t-t_r)\big] + u(t-t_r)$$

The weak-coupling approximations are more closely satisfied for the left case. Eqs. (11.2.7) predict for $V_2(0,t)$ a trapezoidal pulse of duration $2T$ and height K_b, and for $V_2(l,t)$, a rectangular pulse of width t_r and height $K_f/t_r = -0.2$ starting at $t = T$:

$$V_2(l,t) = K_f\frac{dV(t-T)}{dt} = \frac{K_f}{t_r}\big[u(t-T) - u(t-T-t_r)\big]$$

These predictions are approximately correct as can be seen in the figure. The approximation predicts also that $V_1(0,t) = V(t)$ and $V_1(l,t) = V(t-T)$, which are not quite true—the effect of line-2 on line-1 cannot be ignored completely.

The interaction between the two lines is seen better in the MATLAB movie `xtalkmovie.m`, which plots the waves $V_1(z,t)$ and $V_2(z,t)$ as they propagate to and get reflected from their respective loads, and compares them to the uncoupled case $V_0(z,t) = V(t - z/v_0)$. The waves $V_{1,2}(z,t)$ are computed by the same method as for the movie `pulsemovie.m` of Example 10.15.1, applied separately to the even and odd modes. □

11.3 Weakly Coupled Lines with Arbitrary Terminations

The even-odd mode decomposition can be carried out only in the case of *identical* lines both of which have the *same* load and generator impedances. The case of arbitrary terminations has been solved in closed form only for homogeneous media [887,890]. It has also been solved for arbitrary media under the weak coupling assumption [897].

Following [897], we solve the general equations (11.1.7)–(11.1.9) for weakly coupled lines assuming arbitrary terminating impedances Z_{Li}, Z_{Gi}, with reflection coefficients:

$$\Gamma_{Li} = \frac{Z_{Li} - Z_i}{Z_{Li} + Z_i}, \quad \Gamma_{Gi} = \frac{Z_{Gi} - Z_i}{Z_{Gi} + Z_i}, \quad i = 1, 2 \tag{11.3.1}$$

Working with the forward and backward waves, we write Eq. (11.1.7) as the 4×4 matrix equation:

$$\frac{d\mathbf{c}}{dz} = -jM\mathbf{c}, \qquad \mathbf{c} = \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}, \quad M = \begin{bmatrix} \beta_1 & \kappa & 0 & -\chi \\ \kappa & \beta_2 & -\chi & 0 \\ 0 & \chi & -\beta_1 & -\kappa \\ \chi & 0 & -\kappa & -\beta_2 \end{bmatrix}$$

The weak coupling assumption consists of ignoring the coupling of a_1, b_1 on a_2, b_2. This amounts to approximating the above linear system by:

$$\frac{d\mathbf{c}}{dz} = -j\hat{M}\mathbf{c}, \qquad \hat{M} = \begin{bmatrix} \beta_1 & 0 & 0 & 0 \\ \kappa & \beta_2 & -\chi & 0 \\ 0 & 0 & -\beta_1 & 0 \\ \chi & 0 & -\kappa & -\beta_2 \end{bmatrix} \tag{11.3.2}$$

Its solution is given by $\mathbf{c}(z) = e^{-j\hat{M}z}\mathbf{c}(0)$, where the transition matrix $e^{-j\hat{M}z}$ can be expressed in closed form as follows:

$$e^{-j\hat{M}z} = \begin{bmatrix} e^{-j\beta_1 z} & 0 & 0 & 0 \\ \hat{\kappa}(e^{-j\beta_1 z} - e^{-j\beta_2 z}) & e^{-j\beta_2 z} & \hat{\chi}(e^{j\beta_1 z} - e^{-j\beta_2 z}) & 0 \\ 0 & 0 & e^{j\beta_1 z} & 0 \\ \hat{\chi}(e^{-j\beta_1 z} - e^{j\beta_2 z}) & 0 & \hat{\kappa}(e^{j\beta_1 z} - e^{j\beta_2 z}) & e^{j\beta_2 z} \end{bmatrix}, \qquad \begin{aligned} \hat{\kappa} &= \frac{\kappa}{\beta_1 - \beta_2} \\ \hat{\chi} &= \frac{\chi}{\beta_1 + \beta_2} \end{aligned}$$

The transition matrix $e^{-j\hat{M}l}$ may be written in terms of the z-domain delay variables $\zeta_i = e^{j\beta_i l} = e^{j\omega T_i}$, $i = 1, 2$, where T_i are the one-way travel times along the lines, that is, $T_i = l/v_i$. Then, we find:

$$\begin{bmatrix} a_1(l) \\ a_2(l) \\ b_1(l) \\ b_2(l) \end{bmatrix} = \begin{bmatrix} \zeta_1^{-1} & 0 & 0 & 0 \\ \hat{\kappa}(\zeta_1^{-1} - \zeta_2^{-1}) & \zeta_2^{-1} & \hat{\chi}(\zeta_1 - \zeta_2^{-1}) & 0 \\ 0 & 0 & \zeta_1 & 0 \\ \hat{\chi}(\zeta_1^{-1} - \zeta_2) & 0 & \hat{\kappa}(\zeta_1 - \zeta_2) & \zeta_2 \end{bmatrix} \begin{bmatrix} a_1(0) \\ a_2(0) \\ b_1(0) \\ b_2(0) \end{bmatrix} \tag{11.3.3}$$

These must be appended by the appropriate terminating conditions. Assuming that only line-1 is driven, we have:

$$\begin{aligned} &V_1(0)+Z_{G1}I_1(0)=V_{G1}\,, \quad && V_1(l)=Z_{L1}I_1(l) \\ &V_2(0)+Z_{G2}I_2(0)=0\,, && V_2(l)=Z_{L2}I_2(l) \end{aligned}$$

which can be written in terms of the $\mathbf{a},\mathbf{b}$ waves:

$$\begin{aligned} &a_1(0)-\Gamma_{G1}b_1(0)=U_1\,, \quad && b_1(l)=\Gamma_{L1}a_1(l) \\ &a_2(0)-\Gamma_{G2}b_2(0)=0\,, && b_2(l)=\Gamma_{L2}a_2(l) \end{aligned}\;, \quad U_1=\sqrt{\frac{2}{Z_1}}(1-\Gamma_{G1})\frac{V_{G1}}{2} \tag{11.3.4}$$

Eqs. (11.3.3) and (11.3.4) provide a set of eight equations in eight unknowns. Once these are solved, the near- and far-end voltages may be determined. For line-1, we find:

$$\begin{aligned} V_1(0) &= \sqrt{\frac{Z_1}{2}}\big[a_1(0)+b_1(0)\big] = \frac{1+\Gamma_{L1}\zeta_1^{-2}}{1-\Gamma_{G1}\Gamma_{L1}\zeta_1^{-2}}\,V \\ V_1(l) &= \sqrt{\frac{Z_1}{2}}\big[a_1(l)+b_1(l)\big] = \frac{\zeta_1^{-1}(1+\Gamma_{L1})}{1-\Gamma_{G1}\Gamma_{L1}\zeta_1^{-2}}\,V \end{aligned} \tag{11.3.5}$$

where $V = (1-\Gamma_{G1})V_{G1}/2 = Z_1V_{G1}/(Z_1+Z_{G1})$. For line-2, we have:

$$\begin{aligned} V_2(0) &= \frac{\bar{\kappa}(\zeta_1^{-1}-\zeta_2^{-1})(\Gamma_{L1}\zeta_1^{-1}+\Gamma_{L2}\zeta_2^{-1})+\bar{\chi}(1-\zeta_1^{-1}\zeta_2^{-1})(1+\Gamma_{L1}\Gamma_{L2}\zeta_1^{-1}\zeta_2^{-1})}{(1-\Gamma_{G1}\Gamma_{L1}\zeta_1^{-2})(1-\Gamma_{G2}\Gamma_{L2}\zeta_2^{-2})}\,V_{20} \\ V_2(l) &= \frac{\bar{\kappa}(\zeta_1^{-1}-\zeta_2^{-1})(1+\Gamma_{L1}\Gamma_{G2}\zeta_1^{-1}\zeta_2^{-1})+\bar{\chi}(1-\zeta_1^{-1}\zeta_2^{-1})(\Gamma_{L1}\zeta_1^{-1}+\Gamma_{G2}\zeta_2^{-1})}{(1-\Gamma_{G1}\Gamma_{L1}\zeta_1^{-2})(1-\Gamma_{G2}\Gamma_{L2}\zeta_2^{-2})}\,V_{2l} \end{aligned} \tag{11.3.6}$$

where $V_{20} = (1+\Gamma_{G2})V = (1+\Gamma_{G2})(1-\Gamma_{G1})V_{G1}/2$ and $V_{2l} = (1+\Gamma_{L2})V$, and we defined $\bar{\kappa},\bar{\chi}$ by:

$$\begin{aligned} \bar{\kappa} &= \sqrt{\frac{Z_2}{Z_1}}\,\hat{\kappa} = \sqrt{\frac{Z_2}{Z_1}}\,\frac{\kappa}{\beta_1-\beta_2} = \frac{\omega}{\beta_1-\beta_2}\,\frac{1}{2}\left(\frac{L_m}{Z_1}-C_mZ_2\right) \\ \bar{\chi} &= \sqrt{\frac{Z_2}{Z_1}}\,\hat{\chi} = \sqrt{\frac{Z_2}{Z_1}}\,\frac{\chi}{\beta_1+\beta_2} = \frac{\omega}{\beta_1+\beta_2}\,\frac{1}{2}\left(\frac{L_m}{Z_1}+C_mZ_2\right) \end{aligned} \tag{11.3.7}$$

In the case of identical lines with $Z_1 = Z_2 = Z_0$ and $\beta_1 = \beta_2 = \beta = \omega/v_0$, we must take the limit:

$$\lim_{\beta_2\to\beta_1}\frac{e^{-j\beta_1 l}-e^{-j\beta_2 l}}{\beta_1-\beta_2} = \frac{d}{d\beta_1}e^{-j\beta_1 l} = -jle^{-j\beta_1 l}$$

Then, we obtain:

$$\begin{aligned} &\bar{\kappa}(\zeta_1^{-1}-\zeta_2^{-1}) \to j\omega K_f e^{-j\beta l} = -j\omega\frac{l}{2}\left(\frac{L_m}{Z_0}-C_mZ_0\right)e^{-j\beta l} \\ &\bar{\chi} \to K_b = \frac{v_0}{4}\left(\frac{L_m}{Z_0}+C_mZ_0\right) \end{aligned} \tag{11.3.8}$$

where K_f, K_b were defined in (11.2.8). Setting $\zeta_1 = \zeta_2 = \zeta = e^{j\beta l} = e^{j\omega T}$, we obtain the crosstalk signals:

$$\begin{aligned} V_2(0) &= \frac{j\omega K_f(\Gamma_{L1}+\Gamma_{L2})\zeta^{-2}+K_b(1-\zeta^{-2})(1+\Gamma_{L1}\Gamma_{L2}\zeta^{-2})}{(1-\Gamma_{G1}\Gamma_{L1}\zeta^{-2})(1-\Gamma_{G2}\Gamma_{L2}\zeta^{-2})}\,V_{20} \\ V_2(l) &= \frac{j\omega K_f(1+\Gamma_{L1}\Gamma_{G2}\zeta^{-2})\zeta^{-1}+K_b(1-\zeta^{-2})(\Gamma_{L1}+\Gamma_{G2})\zeta^{-1}}{(1-\Gamma_{G1}\Gamma_{L1}\zeta^{-2})(1-\Gamma_{G2}\Gamma_{L2}\zeta^{-2})}\,V_{2l} \end{aligned} \tag{11.3.9}$$

The corresponding time-domain signals will involve the double multiple reflections arising from the denominators. However, if we assume the each line is matched in at least one of its ends, so that $\Gamma_{G1}\Gamma_{L1} = \Gamma_{G2}\Gamma_{L2} = 0$, then the denominators can be eliminated. Replacing $j\omega$ by the time-derivative d/dt and each factor ζ^{-1} by a delay by T, we obtain:

$$\begin{aligned} V_2(0,t) &= K_f(\Gamma_{L1}+\Gamma_{L2}+\Gamma_{L1}\Gamma_{G2})\dot{V}(t-2T) \\ &\quad + K_b(1+\Gamma_{G2})\big[V(t)-V(t-2T)\big] + K_b\Gamma_{L1}\Gamma_{L2}\big[V(t-2T)-V(t-4T)\big] \\ V_2(l,t) &= K_f\big[(1+\Gamma_{L2})\dot{V}(t-T)+\Gamma_{L1}\Gamma_{G2}\dot{V}(t-3T)\big] \\ &\quad + K_b(\Gamma_{L1}+\Gamma_{G2}+\Gamma_{L1}\Gamma_{L2})\big[V(t-T)-V(t-3T)\big] \end{aligned} \tag{11.3.10}$$

where $V(t) = (1-\Gamma_{G1})V_{G1}(t)/2$, and we used the property $\Gamma_{G2}\Gamma_{L2} = 0$ to simplify the expressions. Eqs. (11.3.10) reduce to (11.2.7) when the lines are matched at both ends.

11.4 Coupled-Mode Theory

In its simplest form, *coupled-mode* or *coupled-wave theory* provides a paradigm for the interaction between two waves and the exchange of energy from one to the other as they propagate. Reviews and earlier literature may be found in Refs. [898–919], see also [747–766] for the relationship to fiber Bragg gratings and distributed feedback lasers.

There are several mechanical and electrical analogs of coupled-mode theory, such as a pair of coupled pendula, or two masses at the ends of two springs with a third spring connecting the two, or two LC circuits with a coupling capacitor between them. In these examples, the exchange of energy is taking place over time instead of over space.

Coupled-wave theory is inherently directional. If two forward-moving waves are strongly coupled, then their interactions with the corresponding backward waves may be ignored. Similarly, if a forward- and a backward-moving wave are strongly coupled, then their interactions with the corresponding oppositely moving waves may be ignored. Fig. 11.4.1 depicts these two cases of *co-directional* and *contra-directional* coupling.

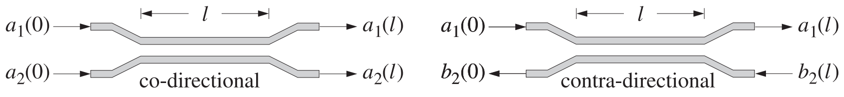


Fig. 11.4.1 Directional Couplers.

Eqs. (11.1.7) form the basis of coupled-mode theory. In the co-directional case, if we assume that there are only forward waves at $z = 0$, that is, $\mathbf{a}(0) \neq 0$ and $\mathbf{b}(0) = 0$, then it may shown that the effect of the backward waves on the forward ones becomes a second-order effect in the coupling constants, and therefore, it may be ignored. To see this, we solve the second of Eqs. (11.1.7) for **b** in terms of **a**, assuming zero initial conditions, and substitute it in the first:

$$\mathbf{b}(z)=-j\int_0^z e^{jF(z-z')}G\,\mathbf{a}(z')\,dz' \quad\Rightarrow\quad \frac{d\mathbf{a}}{dz}=-jF\,\mathbf{a}+\int_0^z Ge^{jF(z-z')}G\,\mathbf{a}(z')\,dz'$$

The second term is second-order in G, or in the coupling constant χ. Ignoring this term, we obtain the standard equations describing a co-directional coupler:

$$\frac{d\mathbf{a}}{dz}=-jF\,\mathbf{a} \quad\Rightarrow\quad \frac{d}{dz}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=-j\begin{bmatrix} \beta_1 & \kappa \\ \kappa & \beta_2 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{11.4.1}$$

For the contra-directional case, a similar argument that assumes the initial conditions $a_2(0)=b_1(0)=0$ gives the following approximation that couples the a_1 and b_2 waves:

$$\frac{d}{dz}\begin{bmatrix} a_1 \\ b_2 \end{bmatrix}=-j\begin{bmatrix} \beta_1 & -\chi \\ \chi & -\beta_2 \end{bmatrix}\begin{bmatrix} a_1 \\ b_2 \end{bmatrix} \tag{11.4.2}$$

The conserved powers are in the two cases:

$$P=|a_1|^2+|a_2|^2\,,\qquad P=|a_1|^2-|b_2|^2 \tag{11.4.3}$$

The solution of Eq. (11.4.1) is obtained with the help of the transition matrix e^{-jFz}:

$$e^{-jFz}=e^{-j\beta z}\begin{bmatrix} \cos\sigma z-j\dfrac{\delta}{\sigma}\sin\sigma z & -j\dfrac{\kappa}{\sigma}\sin\sigma z \\ -j\dfrac{\kappa}{\sigma}\sin\sigma z & \cos\sigma z+j\dfrac{\delta}{\sigma}\sin\sigma z \end{bmatrix} \tag{11.4.4}$$

where

$$\beta=\frac{\beta_1+\beta_2}{2}\,,\quad \delta=\frac{\beta_1-\beta_2}{2}\,,\quad \sigma=\sqrt{\delta^2+\kappa^2} \tag{11.4.5}$$

Thus, the solution of (11.4.1) is:

$$\begin{bmatrix} a_1(z) \\ a_2(z) \end{bmatrix}=e^{-j\beta z}\begin{bmatrix} \cos\sigma z-j\dfrac{\delta}{\sigma}\sin\sigma z & -j\dfrac{\kappa}{\sigma}\sin\sigma z \\ -j\dfrac{\kappa}{\sigma}\sin\sigma z & \cos\sigma z-j\dfrac{\delta}{\sigma}\sin\sigma z \end{bmatrix}\begin{bmatrix} a_1(0) \\ a_2(0) \end{bmatrix} \tag{11.4.6}$$

Starting with initial conditions $a_1(0)=1$ and $a_2(0)=0$, the total initial power will be $P=|a_1(0)|^2+|a_2(0)|^2=1$. As the waves propagate along the z-direction, power is exchanged between lines 1 and 2 according to:

$$\begin{aligned} P_1(z)&=|a_1(z)|^2=\cos^2\sigma z+\frac{\delta^2}{\sigma^2}\sin^2\sigma z \\ P_2(z)&=|a_2(z)|^2=\frac{\kappa^2}{\sigma^2}\sin^2\sigma z=1-P_1(z) \end{aligned} \tag{11.4.7}$$

Fig. 11.4.2 shows the two cases for which $\delta/\kappa=0$ and $\delta/\kappa=0.5$. In both cases, maximum exchange of power occurs periodically at distances that are odd multiples of $z=\pi/2\sigma$. Complete power exchange occurs only in the case $\delta=0$, or equivalently, when $\beta_1=\beta_2$. In this case, we have $\sigma=\kappa$ and $P_1(z)=\cos^2\kappa z$, $P_2(z)=\sin^2\kappa z$.

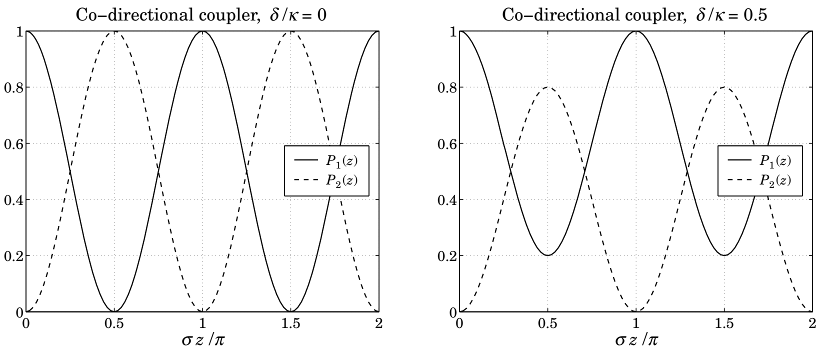


Fig. 11.4.2 Power exchange in co-directional couplers.

11.5 *Fiber Bragg Gratings*

As an example of contra-directional coupling, we consider the case of a *fiber Bragg grating* (FBG), that is, a fiber with a segment that has a periodically varying refractive index, as shown in Fig. 11.5.1.

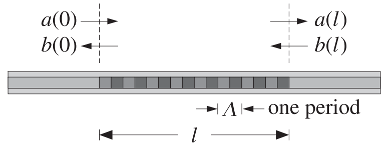


Fig. 11.5.1 Fiber Bragg grating.

The backward wave is generated by the reflection of a forward-moving wave incident on the interface from the left. The grating behaves very similarly to a periodic multilayer structure, such as a dielectric mirror at normal incidence, exhibiting high-reflectance bands. A simple model for an FBG is as follows [747–766]:

$$\frac{d}{dz}\begin{bmatrix} a(z) \\ b(z) \end{bmatrix}=-j\begin{bmatrix} \beta & \kappa e^{-jKz} \\ -\kappa^* e^{jKz} & -\beta \end{bmatrix}\begin{bmatrix} a(z) \\ b(z) \end{bmatrix} \tag{11.5.1}$$

where $K=2\pi/\Lambda$ is the Bloch wavenumber, Λ is the period, and $a(z),b(z)$ represent the forward and backward waves. The following transformation removes the phase factor e^{-jKz} from the coupling constant:

$$\begin{bmatrix} A(z) \\ B(z) \end{bmatrix}=\begin{bmatrix} e^{jKz/2} & 0 \\ 0 & e^{-jKz/2} \end{bmatrix}\begin{bmatrix} a(z) \\ b(z) \end{bmatrix}=\begin{bmatrix} e^{jKz/2}a(z) \\ e^{-jKz/2}b(z) \end{bmatrix} \tag{11.5.2}$$

$$\frac{d}{dz}\begin{bmatrix} A(z) \\ B(z) \end{bmatrix}=-j\begin{bmatrix} \delta & \kappa \\ -\kappa^* & -\delta \end{bmatrix}\begin{bmatrix} A(z) \\ B(z) \end{bmatrix} \tag{11.5.3}$$

where $\delta = \beta - K/2$ is referred to as a *detuning parameter*. The conserved power is given by $P(z) = |a(z)|^2 - |b(z)|^2$. The fields at $z = 0$ are related to those at $z = l$ by:

$$\begin{bmatrix} A(0) \\ B(0) \end{bmatrix} = e^{jFl} \begin{bmatrix} A(l) \\ B(l) \end{bmatrix}, \quad \text{with} \quad F = \begin{bmatrix} \delta & \kappa \\ -\kappa^* & -\delta \end{bmatrix} \tag{11.5.4}$$

The transfer matrix e^{jFl} is given by:

$$e^{jFl} = \begin{bmatrix} \cos\sigma l + j\dfrac{\delta}{\sigma}\sin\sigma l & j\dfrac{\kappa}{\sigma}\sin\sigma l \\ -j\dfrac{\kappa^*}{\sigma}\sin\sigma l & \cos\sigma l - j\dfrac{\delta}{\sigma}\sin\sigma l \end{bmatrix} \equiv \begin{bmatrix} U_{11} & U_{12} \\ U_{12}^* & U_{11}^* \end{bmatrix} \tag{11.5.5}$$

where $\sigma = \sqrt{\delta^2 - |\kappa|^2}$. If $|\delta| < |\kappa|$, then σ becomes imaginary. In this case, it is more convenient to express the transfer matrix in terms of the quantity $\gamma = \sqrt{|\kappa|^2 - \delta^2}$:

$$e^{jFl} = \begin{bmatrix} \cosh\gamma l + j\dfrac{\delta}{\gamma}\sinh\gamma l & j\dfrac{\kappa}{\gamma}\sinh\gamma l \\ -j\dfrac{\kappa^*}{\gamma}\sinh\gamma l & \cosh\gamma l - j\dfrac{\delta}{\gamma}\sinh\gamma l \end{bmatrix} \tag{11.5.6}$$

The transfer matrix has unit determinant, which implies that $|U_{11}|^2 - |U_{12}|^2 = 1$. Using this property, we may rearrange (11.5.4) into its *scattering matrix* form that relates the outgoing fields to the incoming ones:

$$\begin{bmatrix} B(0) \\ A(l) \end{bmatrix} = \begin{bmatrix} \Gamma & T \\ T & \Gamma' \end{bmatrix} \begin{bmatrix} A(0) \\ B(l) \end{bmatrix}, \quad \Gamma = \frac{U_{12}^*}{U_{11}}, \quad \Gamma' = -\frac{U_{12}}{U_{11}}, \quad T = \frac{1}{U_{11}} \tag{11.5.7}$$

where Γ, Γ' are the *reflection* coefficients from the left and right, respectively, and T is the transmission coefficient. We have explicitly,

$$\Gamma = \frac{-j\dfrac{\kappa^*}{\sigma}\sin\sigma l}{\cos\sigma l + j\dfrac{\delta}{\sigma}\sin\sigma l}, \quad T = \frac{1}{\cos\sigma l + j\dfrac{\delta}{\sigma}\sin\sigma l} \tag{11.5.8}$$

If there is only an incident wave from the left, that is, $A(0) \neq 0$ and $B(l) = 0$, then (11.5.7) implies that $B(0) = \Gamma A(0)$ and $A(l) = TA(0)$.

A consequence of power conservation, $|A(0)|^2 - |B(0)|^2 = |A(l)|^2 - |B(l)|^2$, is the unitarity of the scattering matrix, which implies the property $|\Gamma|^2 + |T|^2 = 1$. The reflectance $|\Gamma|^2$ may be expressed in the following two forms, the first being appropriate when $|\delta| \geq |\kappa|$, and the second when $|\delta| \leq |\kappa|$:

$$|\Gamma|^2 = 1 - |T|^2 = \frac{|\kappa|^2 \sin^2\sigma l}{\sigma^2\cos^2\sigma l + \delta^2\sin^2\sigma l} = \frac{|\kappa|^2\sinh^2\gamma l}{\gamma^2\cosh^2\gamma l + \delta^2\sinh^2\gamma l} \tag{11.5.9}$$

Fig. 11.5.2 shows $|\Gamma|^2$ as a function of δ. The high-reflectance band corresponds to the range $|\delta| \leq |\kappa|$. The left graph has $\kappa l = 3$ and the right one $\kappa l = 6$.

As κl increases, the reflection band becomes sharper. The asymptotic width of the band is $-|\kappa| \leq \delta \leq |\kappa|$. For any finite value of κl, the maximum reflectance achieved

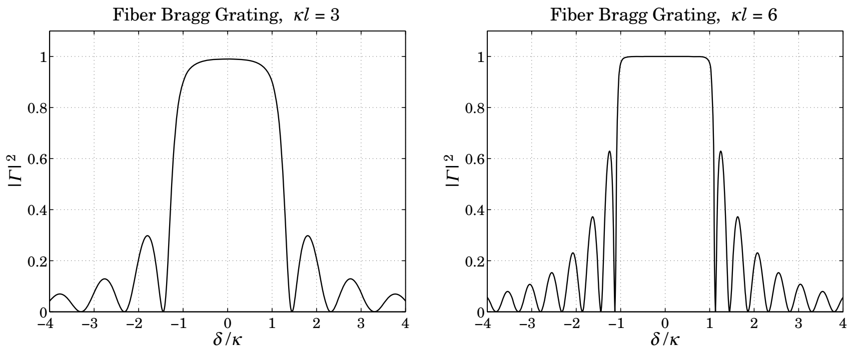


Fig. 11.5.2 Reflectance of fiber Bragg gratings.

at the center of the band, $\delta = 0$, is given by $|\Gamma|^2_{\max} = \tanh^2|\kappa l|$. The reflectance at the asymptotic band edges is given by:

$$|\Gamma|^2 = \frac{|\kappa l|^2}{1 + |\kappa l|^2}, \quad \text{at} \quad \delta = \pm|\kappa|$$

The zeros of the reflectance correspond to $\sin\sigma l = 0$, or, $\sigma = m\pi/l$, which gives $\delta = \pm\sqrt{|\kappa|^2 + (m\pi/l)^2}$, where m is a non-zero integer.

The *Bragg wavelength* λ_B is the wavelength at the center of the reflecting band, that is, corresponding to $\delta = 0$, or, $\beta = K/2$, or $\lambda_B = 2\pi/\beta = 4\pi/K = 2\Lambda$.

By concatenating two identical FBGs separated by a "spacer" of length $d = \lambda_B/4 = \Lambda/2$, we obtain a *quarter-wave phase-shifted FBG*, which has a narrow transmission window centered at $\delta = 0$. Fig. 11.5.3 depicts such a compound grating. Within the spacer, the A, B waves propagate with wavenumber β as though they are uncoupled.

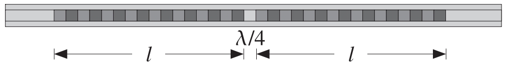


Fig. 11.5.3 Quarter-wave phase-shifted fiber Bragg grating.

The compound transfer matrix is obtained by multiplying the transfer matrices of the two FBGs and the spacer: $V = U_{\text{FBG}}U_{\text{spacer}}U_{\text{FBG}}$, or, explicitly:

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{12}^* & V_{11}^* \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ U_{12}^* & U_{11}^* \end{bmatrix} \begin{bmatrix} e^{j\beta d} & 0 \\ 0 & e^{-j\beta d} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{12}^* & U_{11}^* \end{bmatrix} \tag{11.5.10}$$

where the U_{ij} are given in Eq. (11.5.5). It follows that the matrix elements of V are:

$$V_{11} = U_{11}^2 e^{j\beta d} + |U_{12}|^2 e^{-j\beta d}, \quad V_{12} = U_{12}\big(U_{11}e^{j\beta d} + U_{11}^* e^{-j\beta d}\big) \tag{11.5.11}$$

The reflection coefficient of the compound grating will be:

$$\Gamma_{\text{comp}} = \frac{V_{12}^*}{V_{11}} = \frac{U_{12}\big(U_{11}e^{j\beta d} + U_{11}^* e^{-j\beta d}\big)}{U_{11}^2 e^{j\beta d} + |U_{12}|^2 e^{-j\beta d}} = \frac{\Gamma\big(T^* e^{j\beta d} + Te^{-j\beta d}\big)}{T^* e^{j\beta d} + |\Gamma|^2 Te^{-j\beta d}} \tag{11.5.12}$$

where we replaced $U_{12}^* = \Gamma/T$ and $U_{11} = 1/T$. Assuming a quarter-wavelength spacing $d = \lambda_B/4 = \Lambda/2$, we have $\beta d = (\delta + \pi/\Lambda)d = \delta d + \pi/2$. Replacing $e^{j\beta d} = e^{j\delta d + j\pi/2} = j\,e^{j\delta d}$, we obtain:

$$\Gamma_{\text{comp}} = \frac{\Gamma(T^* e^{j\delta d} - T e^{-j\delta d})}{T^* e^{j\delta d} - |\Gamma|^2 T e^{-j\delta d}} \tag{11.5.13}$$

At $\delta = 0$, we have $T = T^* = 1/\cosh|\kappa|l$, and therefore, $\Gamma_{\text{comp}} = 0$. Fig. 11.5.4 depicts the reflectance, $|\Gamma_{\text{comp}}|^2$, and transmittance, $1 - |\Gamma_{\text{comp}}|^2$, for the case $\kappa l = 2$.

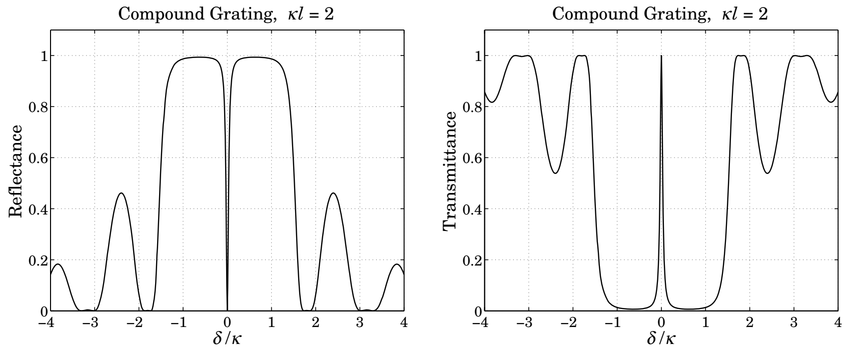


Fig. 11.5.4 Quarter-wave phase-shifted fiber Bragg grating.

Quarter-wave phase-shifted FBGs are similar to the Fabry-Perot resonators discussed in Sec. 6.5. Improved designs having narrow and flat transmission bands can be obtained by cascading several quarter-wave FBGs with different lengths [747–767]. Some applications of FBGs in DWDM systems were pointed out in Sec. 6.7.

11.6 Diffuse Reflection and Transmission

Another example of contra-directional coupling is the two-flux model of Schuster and Kubelka-Munk describing the absorption and multiple scattering of light propagating in a turbid medium [920–936].

The model has a large number of applications, such as radiative transfer in stellar atmospheres, reflectance spectroscopy, reflection and transmission properties of powders, papers, paints, skin tissue, dental materials, and the sea.

The model assumes a simplified parallel-plane geometry, as shown in Fig. 11.6.1. Let $I_\pm(z)$ be the forward and backward radiation intensities per unit frequency interval at location z within the material. The model is described by the two coefficients k, s of absorption and scattering per unit length. For simplicity, we assume that k, s are independent of z.

Within a layer dz, the forward intensity I_+ will be diminished by an amount of $I_+ k\,dz$ due to absorption and an amount of $I_+ s\,dz$ due to scattering, and it will be increased by an amount of $I_- s\,dz$ arising from the backward-moving intensity that is getting scattered forward. Similarly, the backward intensity, going from $z + dz$ to z, will be decreased by $I_-(k+s)(-dz)$ and increased by $I_+ s(-dz)$. Thus, the incremental changes are:

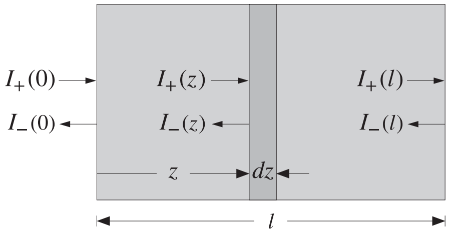


Fig. 11.6.1 Forward and backward intensities in stratified medium.

$$dI_+ = -(k+s)I_+ dz + sI_- dz$$
$$-dI_- = -(k+s)I_- dz + sI_+ dz$$

or, written in matrix form:

$$\frac{d}{dz}\begin{bmatrix} I_+(z) \\ I_-(z) \end{bmatrix} = -\begin{bmatrix} k+s & -s \\ s & -k-s \end{bmatrix}\begin{bmatrix} I_+(z) \\ I_-(z) \end{bmatrix} \tag{11.6.1}$$

This is similar in structure to Eq. (11.5.3), except the matrix coefficients are real. The solution at distance $z = l$ is obtained in terms of the initial values $I_\pm(0)$ by:

$$\begin{bmatrix} I_+(l) \\ I_-(l) \end{bmatrix} = e^{-Fl}\begin{bmatrix} I_+(0) \\ I_-(0) \end{bmatrix}, \quad \text{with} \quad F = \begin{bmatrix} k+s & -s \\ s & -k-s \end{bmatrix} \tag{11.6.2}$$

The transfer matrix e^{-Fl} is:

$$U = e^{-Fl} = \begin{bmatrix} \cosh\beta l - \dfrac{\alpha}{\beta}\sinh\beta l & \dfrac{s}{\beta}\sinh\beta l \\ -\dfrac{s}{\beta}\sinh\beta l & \cosh\beta l + \dfrac{\alpha}{\beta}\sinh\beta l \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \tag{11.6.3}$$

where $\alpha = k + s$ and $\beta = \sqrt{\alpha^2 - s^2} = \sqrt{k(k+2s)}$.[†] The transfer matrix is unimodular, that is, $\det U = U_{11}U_{22} - U_{12}U_{21} = 1$.

Of interest are the input reflectance (the albedo) $R = I_-(0)/I_+(0)$ of the length-l structure and its transmittance $T = I_+(l)/I_+(0)$, both expressed in terms of the output, or background, reflectance $R_g = I_-(l)/I_+(l)$. Using Eq. (11.6.2), we find:

$$\begin{aligned} R &= \frac{-U_{21} + U_{11}R_g}{U_{22} - U_{12}R_g} = \frac{s\sinh\beta l + (\beta\cosh\beta l - \alpha\sinh\beta l)R_g}{\beta\cosh\beta l + (\alpha - sR_g)\sinh\beta l} \\ T &= \frac{1}{U_{22} - U_{12}R_g} = \frac{\beta}{\beta\cosh\beta l + (\alpha - sR_g)\sinh\beta l} \end{aligned} \tag{11.6.4}$$

[†] These are related to the normalized Kubelka [926] variables $a = \alpha/s$, $b = \beta/s$.

The reflectance and transmittance corresponding to a black, non-reflecting, background are obtained by setting $R_g = 0$ in Eq. (11.6.4):

$$R_0 = \frac{-U_{21}}{U_{22}} = \frac{s\sinh\beta l}{\beta\cosh\beta l + \alpha\sinh\beta l}$$
$$T_0 = \frac{1}{U_{22}} = \frac{\beta}{\beta\cosh\beta l + \alpha\sinh\beta l} \tag{11.6.5}$$

The reflectance of an infinitely-thick medium is obtained in the limit $l \to \infty$:

$$R_\infty = \frac{s}{\alpha+\beta} = \frac{s}{k+s+\sqrt{k(k+2s)}} \quad\Rightarrow\quad \frac{k}{s} = \frac{(R_\infty - 1)^2}{2R_\infty} \tag{11.6.6}$$

For the special case of an absorbing but non-scattering medium ($k \neq 0, s = 0$), we have $\alpha = \beta = k$ and the transfer matrix (11.6.3) and Eq. (11.6.4) simplify into:

$$U = e^{-Fl} = \begin{bmatrix} e^{-kl} & 0 \\ 0 & e^{kl} \end{bmatrix}, \quad R = e^{-2kl}R_g\,, \quad T = e^{-kl} \tag{11.6.7}$$

These are in accordance with our expectations for exponential attenuation with distance. The intensities are related by $I_+(l) = e^{-kl}I_+(0)$ and $I_-(l) = e^{kl}I_-(0)$. Thus, the reflectance corresponds to traversing a forward and a reverse path of length l, and the transmittance only a forward path.

Perhaps, the most surprising prediction of this model (first pointed out by Schuster) is that, in the case of a non-absorbing but scattering medium ($k = 0, s \neq 0$), the transmittance is not attenuating exponentially, but rather, inversely with distance. Indeed, setting $\alpha = s$ and taking the limit $\beta^{-1}\sinh\beta l \to l$ as $\beta \to 0$, we find:

$$U = e^{-Fl} = \begin{bmatrix} 1-sl & sl \\ -sl & 1+sl \end{bmatrix}, \quad R = \frac{sl + (1-sl)R_g}{1+sl-slR_g}\,, \quad T = \frac{1}{1+sl-slR_g} \tag{11.6.8}$$

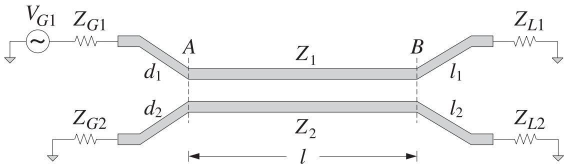
In particular, for the case of a non-reflecting background, we have:

$$R_0 = \frac{sl}{1+sl}\,, \quad T_0 = \frac{1}{1+sl} \tag{11.6.9}$$

11.7 Problems

11.1 Show that the coupled telegrapher's equations (11.1.4) can be written in the form (11.1.7).

11.2 Consider the practical case in which two lines are coupled only over a middle portion of length l, with their beginning and ending segments being uncoupled, as shown below:



Assuming weakly coupled lines, how should Eqs. (11.3.6) and (11.3.9) be modified in this case? [*Hint:* Replace the segments to the left of the reference plane A and to the right of plane B by their Thévenin equivalents.]

11.3 Derive the transition matrix $e^{-j\hat{M}z}$ of weakly coupled lines described by Eq. (11.3.2).

11.4 Verify explicitly that Eq. (11.4.6) is the solution of the coupled-mode equations (11.4.1).

11.5 *Computer Experiment—Fiber Bragg Gratings.* Reproduce the results and graphs of Figures 11.5.2 and 11.5.3.

12

Impedance Matching

12.1 Conjugate and Reflectionless Matching

The Thévenin equivalent circuits depicted in Figs. 10.11.1 and 10.11.3 also allow us to answer the question of *maximum power transfer.* Given a generator and a length-d transmission line, maximum transfer of power from the generator to the load takes place when the load is *conjugate matched* to the generator, that is,

$$\boxed{Z_L = Z_{\text{th}}^*} \qquad \text{(conjugate match)} \tag{12.1.1}$$

The proof of this result is postponed until Sec. 15.4. Writing $Z_{\text{th}} = R_{\text{th}} + jX_{\text{th}}$ and $Z_L = R_L + jX_L$, the condition is equivalent to $R_L = R_{\text{th}}$ and $X_L = -X_{\text{th}}$. In this case, half of the generated power is delivered to the load and half is dissipated in the generator's Thévenin resistance. From the Thévenin circuit shown in Fig. 10.11.1, we find for the current through the load:

$$I_L = \frac{V_{\text{th}}}{Z_{\text{th}} + Z_L} = \frac{V_{\text{th}}}{(R_{\text{th}} + R_L) + j(X_{\text{th}} + X_L)} = \frac{V_{\text{th}}}{2R_{\text{th}}}$$

Thus, the total reactance of the circuit is canceled. It follows then that the power delivered by the Thévenin generator and the powers dissipated in the generator's Thévenin resistance and the load will be:

$$\begin{gathered}
P_{\text{tot}} = \frac{1}{2}\text{Re}(V_{\text{th}}^* I_L) = \frac{|V_{\text{th}}|^2}{4R_{\text{th}}} \\
P_{\text{th}} = \frac{1}{2}R_{\text{th}}|I_L|^2 = \frac{|V_{\text{th}}|^2}{8R_{\text{th}}} = \frac{1}{2}P_{\text{tot}}\,, \quad P_L = \frac{1}{2}R_L|I_L|^2 = \frac{|V_{\text{th}}|^2}{8R_{\text{th}}} = \frac{1}{2}P_{\text{tot}}
\end{gathered} \tag{12.1.2}$$

Assuming a lossless line (real-valued Z_0 and β), the conjugate match condition can also be written in terms of the reflection coefficients corresponding to Z_L and Z_{th}:

$$\boxed{\Gamma_L = \Gamma_{\text{th}}^* = \Gamma_G^* e^{2j\beta d}} \qquad \text{(conjugate match)} \tag{12.1.3}$$

Moving the phase exponential to the left, we note that the conjugate match condition can be written in terms of the same quantities at the *input* side of the transmission line:

$$\boxed{\Gamma_d = \Gamma_L e^{-2j\beta d} = \Gamma_G^*} \quad \Leftrightarrow \quad \boxed{Z_d = Z_G^*} \qquad \text{(conjugate match)} \qquad (12.1.4)$$

Thus, the conjugate match condition can be phrased in terms of the input quantities and the equivalent circuit of Fig. 10.9.1. More generally, there is a conjugate match at *every* point along the line.

Indeed, the line can be cut at any distance l from the load and its entire left segment including the generator can be replaced by a Thévenin-equivalent circuit. The conjugate matching condition is obtained by propagating Eq. (12.1.3) to the left by a distance l, or equivalently, Eq. (12.1.4) to the right by distance $d - l$:

$$\boxed{\Gamma_l = \Gamma_L e^{-2j\beta l} = \Gamma_G^* e^{2j\beta(d-l)}} \qquad \text{(conjugate match)} \qquad (12.1.5)$$

Conjugate matching is not the same as *reflectionless matching*, which refers to matching the load to the line impedance, $Z_L = Z_0$, in order to prevent reflections from the load.

In practice, we must use *matching networks* at one or both ends of the transmission line to achieve the desired type of matching. Fig. 12.1.1 shows the two typical situations that arise.

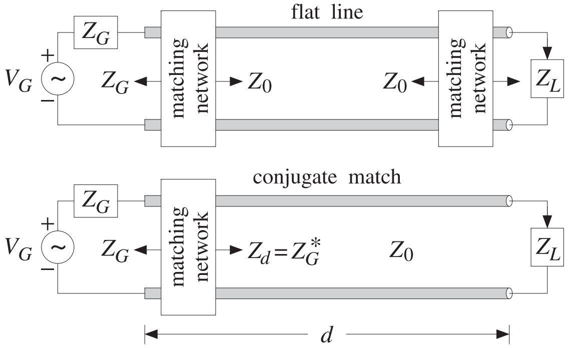


Fig. 12.1.1 Reflectionless and conjugate matching of a transmission line.

In the first, referred to as a *flat line*, both the generator and the load are matched so that effectively, $Z_G = Z_L = Z_0$. There are no reflected waves and the generator (which is typically designed to operate into Z_0) transmits maximum power to the load, as compared to the case when $Z_G = Z_0$ but $Z_L \neq Z_0$.

In the second case, the load is connected to the line without a matching circuit and the generator is conjugate-matched to the input impedance of the line, that is, $Z_d = Z_G^*$. As we mentioned above, the line remains conjugate matched everywhere along its length, and therefore, the matching network can be inserted at any convenient point, not necessarily at the line input.

Because the value of Z_d depends on Z_L and the frequency ω (through $\tan\beta d$), the conjugate match will work as designed only at a single frequency. On the other hand, if the load and generator are purely resistive and are matched individually to the line, the matching will remain reflectionless over a larger frequency bandwidth.

Conjugate matching is usually accomplished using L-section reactive networks. Reflectionless matching is achieved by essentially the same methods as antireflection coating. In the next few sections, we discuss several methods for reflectionless and conjugate matching, such as (a) quarter-wavelength single- and multi-section transformers; (b) two-section series impedance transformers; (c) single, double, and triple stub tuners; and (d) L-section lumped-parameter reactive matching networks.

12.2 Multisection Transmission Lines

Multisection transmission lines are used primarily in the construction of broadband matching terminations. A typical multisection line is shown in Fig. 12.2.1.

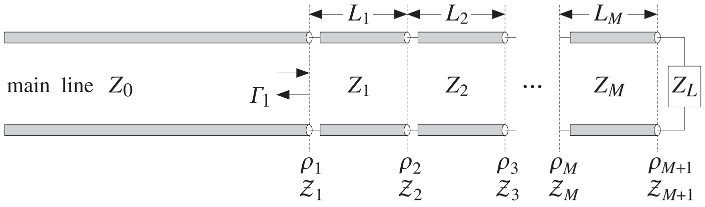


Fig. 12.2.1 Multi-section transmission line.

It consists of M segments between the main line and the load. The ith segment is characterized by its characteristic impedance Z_i, length l_i, and velocity factor, or equivalently, refractive index n_i. The speed in the ith segment is $c_i = c_0/n_i$. The phase thicknesses are defined by:

$$\delta_i = \beta_i l_i = \frac{\omega}{c_i}\, l_i = \frac{\omega}{c_0}\, n_i l_i\,, \quad i = 1, 2, \dots, M \qquad (12.2.1)$$

We may define the electrical lengths (playing the same role as the optical lengths of dielectric slabs) in units of some reference free-space wavelength λ_0 or corresponding frequency $f_0 = c_0/\lambda_0$ as follows:

$$\text{(electrical lengths)} \qquad \boxed{L_i = \frac{n_i l_i}{\lambda_0} = \frac{l_i}{\lambda_i}}\,, \quad i = 1, 2, \dots, M \qquad (12.2.2)$$

where $\lambda_i = \lambda_0/n_i$ is the wavelength within the ith segment. Typically, the electrical lengths are quarter-wavelengths, $L_i = 1/4$. It follows that the phase thicknesses can be expressed in terms of L_i as $\delta_i = \omega n_i l_i/c_0 = 2\pi f n_i l_i/(f_0\lambda_0)$, or,

$$\text{(phase thicknesses)} \qquad \boxed{\delta_i = \beta_i l_i = 2\pi L_i \frac{f}{f_0} = 2\pi L_i \frac{\lambda_0}{\lambda}}\,, \quad i = 1, 2, \dots, M \qquad (12.2.3)$$

where f is the operating frequency and $\lambda = c_0/f$ the corresponding free-space wavelength. The wave impedances, $\mathcal{Z}_i$, are continuous across the $M + 1$ interfaces and are related by the recursions:

$$\mathcal{Z}_i = Z_i\,\frac{\mathcal{Z}_{i+1} + jZ_i\tan\delta_i}{Z_i + j\mathcal{Z}_{i+1}\tan\delta_i}\,,\quad i = M,\dots,1 \tag{12.2.4}$$

and initialized by $\mathcal{Z}_{M+1} = Z_L$. The corresponding reflection responses at the left of each interface, $\Gamma_i = (\mathcal{Z}_i - Z_{i-1})/(\mathcal{Z}_i + Z_{i-1})$, are obtained from the recursions:

$$\boxed{\Gamma_i = \frac{\rho_i + \Gamma_{i+1}e^{-2j\delta_i}}{1 + \rho_i\Gamma_{i+1}e^{-2j\delta_i}}}\,,\quad i = M,\dots,1 \tag{12.2.5}$$

and initialized at $\Gamma_{M+1} = \Gamma_L = (Z_L - Z_M)/(Z_L + Z_M)$, where ρ_i are the elementary reflection coefficients at the interfaces:

$$\rho_i = \frac{Z_i - Z_{i-1}}{Z_i + Z_{i-1}}\,,\quad i = 1,2,\dots,M+1 \tag{12.2.6}$$

where $Z_{M+1} = Z_L$. The MATLAB function `multiline` calculates the reflection response $\Gamma_1(f)$ at interface-1 as a function of frequency. Its usage is:

```
Gamma1 = multiline(Z,L,ZL,f);          % reflection response of multisection line
```

where $Z = [Z_0, Z_1, \dots, Z_M]$ and $L = [L_1, L_2, \dots, L_M]$ are the main line and segment impedances and the segment electrical lengths.

The function `multiline` implements Eq. (12.2.6) and is similar to `multidiel`, except here the load impedance Z_L is a separate input in order to allow it to be a function of frequency. We will see examples of its usage below.

12.3 Quarter-Wavelength Chebyshev Transformers

Quarter-wavelength Chebyshev impedance transformers allow the matching of real-valued load impedances Z_L to real-valued line impedances Z_0 and can be designed to achieve desired attenuation and bandwidth specifications.

The design method has already been discussed in Sec. 6.8. The results of that section translate verbatim to the present case by replacing refractive indices n_i by line admittances $Y_i = 1/Z_i$. Typical design specifications are shown in Fig. 6.8.1.

In an M-section transformer, all segments have equal electrical lengths, $L_i = l_i/\lambda_i = n_il_i/\lambda_0 = 1/4$ at some operating wavelength λ_0. The phase thicknesses of the segments are all equal and are given by $\delta_i = 2\pi L_i f/f_0$, or, because $L_i = 1/4$:

$$\delta = \frac{\pi}{2}\frac{f}{f_0} \tag{12.3.1}$$

The reflection response $|\Gamma_1(f)|^2$ at the left of interface-1 is expressed in terms of the order-M Chebyshev polynomials $T_M(x)$, where x is related to the phase thickness by $x = x_0\cos\delta$:

$$|\Gamma_1(f)|^2 = \frac{e_1^2T_M^2(x_0\cos\delta)}{1 + e_1^2T_M^2(x_0\cos\delta)} \tag{12.3.2}$$

where $e_1 = e_0/T_M(x_0)$ and e_0 is given in terms of the load and main line impedances:

$$e_0^2 = \frac{(Z_L - Z_0)^2}{4Z_LZ_0} = \frac{|\Gamma_L|^2}{1 - |\Gamma_L|^2}\,,\quad \Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{12.3.3}$$

The parameter x_0 is related to the desired reflectionless bandwidth Δf by:

$$x_0 = \frac{1}{\sin\left(\dfrac{\pi}{4}\dfrac{\Delta f}{f_0}\right)} \tag{12.3.4}$$

and $T_M(x_0)$ is related to the attenuation A in the reflectionless band by:

$$A = 10\log_{10}\left(\frac{T_M^2(x_0) + e_0^2}{1 + e_0^2}\right) \tag{12.3.5}$$

Solving for M in terms of A, we have (rounding up to the next integer):

$$M = \text{ceil}\left(\frac{\text{acosh}\left(\sqrt{(1 + e_0^2)\,10^{A/10} - e_0^2}\right)}{\text{acosh}(x_0)}\right) \tag{12.3.6}$$

where A is in dB and is measured from dc, or equivalently, with respect to the reflection response $|\Gamma_L|$ of the unmatched line. The maximum equiripple level within the reflectionless band is given by

$$|\Gamma_1|_{\max} = |\Gamma_L|\,10^{-A/20} \quad\Rightarrow\quad A = 20\log_{10}\left(\frac{|\Gamma_L|}{|\Gamma_1|_{\max}}\right) \tag{12.3.7}$$

This condition can also be expressed in terms of the maximum SWR within the desired bandwidth. Indeed, setting $S_{\max} = (1 + |\Gamma_1|_{\max})/(1 - |\Gamma_1|_{\max})$ and $S_L = (1 + |\Gamma_L|)/(1 - |\Gamma_L|)$, we may rewrite (12.3.7) as follows:

$$\boxed{A = 20\log_{10}\left(\frac{|\Gamma_L|}{|\Gamma_1|_{\max}}\right) = 20\log_{10}\left(\frac{S_L - 1}{S_L + 1}\,\frac{S_{\max} + 1}{S_{\max} - 1}\right)} \tag{12.3.8}$$

where we must demand $S_{\max} < S_L$ or $|\Gamma_1|_{\max} < |\Gamma_L|$. The MATLAB functions `chebtr`, `chebtr2`, and `chebtr3` implement the design steps. In the present context, they have usage:

```
[Y,a,b]    = chebtr(Y0,YL,A,DF);         % Chebyshev multisection transformer design
[Y,a,b,A]  = chebtr2(Y0,YL,M,DF);        % specify order and bandwidth
[Y,a,b,DF] = chebtr3(Y0,YL,M,A);         % specify order and attenuation
```

The outputs are the admittances $\mathbf{Y} = [Y_0, Y_1, Y_2, \dots, Y_M, Y_L]$ and the reflection and transmission polynomials $\mathbf{a}, \mathbf{b}$. In `chebtr2` and `chebtr3`, the order M is given. The designed segment impedances Z_i, $i = 1,2,\dots,M$ satisfy the symmetry properties:

$$Z_iZ_{M+1-i} = Z_0Z_L\,,\quad i = 1,2,\dots,M \tag{12.3.9}$$

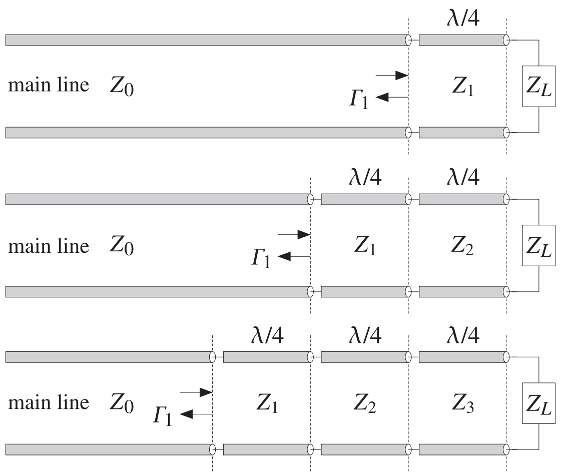


Fig. 12.3.1 One, two, and three-section quarter-wavelength transformers.

Fig. 12.3.1 depicts the three cases of $M = 1, 2, 3$ segments. The case $M = 1$ is used widely and we discuss it in more detail. According to Eq. (12.3.9), the segment impedance satisfies $Z_1^2 = Z_0 Z_L$, or,

$$Z_1 = \sqrt{Z_0 Z_L} \tag{12.3.10}$$

This implies that the reflection coefficients at interfaces 1 and 2 are equal:

$$\rho_1 = \frac{Z_1 - Z_0}{Z_1 + Z_0} = \frac{Z_L - Z_1}{Z_L + Z_1} = \rho_2 \tag{12.3.11}$$

Because the Chebyshev polynomial of order-1 is $T_1(x) = x$, the reflection response (12.3.2) takes the form:

$$|\Gamma_1(f)|^2 = \frac{e_0^2 \cos^2\delta}{1 + e_0^2\cos^2\delta} \tag{12.3.12}$$

Using Eq. (12.3.11), we can easily verify that e_0 is related to ρ_1 by

$$e_0^2 = \frac{4\rho_1^2}{(1-\rho_1^2)^2}$$

Then, Eq. (12.3.12) can be cast in the following equivalent form, which is recognized as the propagation of the load reflection response $\Gamma_2 = \rho_2 = \rho_1$ by a phase thickness δ to interface-1:

$$|\Gamma_1(f)|^2 = \left|\frac{\rho_1(1+z^{-1})}{1+\rho_1^2 z^{-1}}\right|^2 \tag{12.3.13}$$

where $z = e^{2j\delta}$. The reflection response has a zero at $z = -1$ or $\delta = \pi/2$, which occurs at $f = f_0$ and at *odd* multiples of f_0. The wave impedance at interface-1 will be:

$$\mathcal{Z}_1 = Z_1 \frac{Z_L + jZ_1\tan\delta}{Z_0 + jZ_L\tan\delta} \tag{12.3.14}$$

Using Eq. (12.3.10), we obtain the matching condition at $f = f_0$, or at $\delta = \pi/2$:

$$\mathcal{Z}_1 = \frac{Z_1^2}{Z_L} = Z_0 \tag{12.3.15}$$

Example 12.3.1: *Single-section quarter wavelength transformer.* Design a single-section transformer that will match a 200-ohm load to a 50-ohm line at 100 MHz. Determine the bandwidth over which the SWR on the line remains less than 1.5.

Solution: The quarter-wavelength section has impedance $Z_1 = \sqrt{Z_L Z_0} = \sqrt{200\cdot 50} = 100$ ohm. The reflection response $|\Gamma_1(f)|$ and the SWR $S(f) = \big(1+|\Gamma_1(f)|\big)/\big(1-|\Gamma_1(f)|\big)$ are plotted in Fig. 12.3.1 versus frequency.

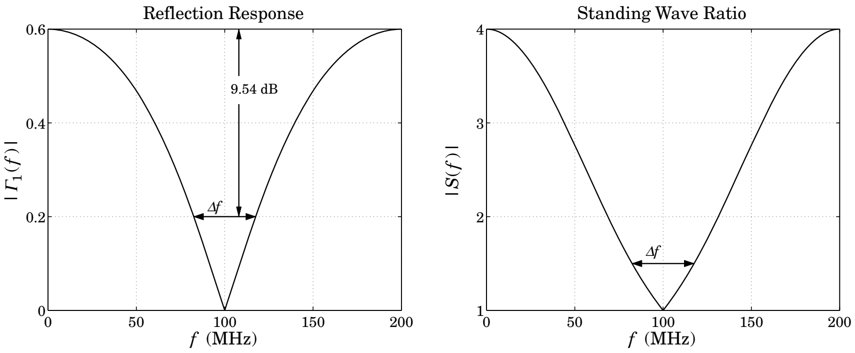


Fig. 12.3.2 Reflection response and line SWR of single-section transformer.

The reflection coefficient of the unmatched line and the maximum tolerable reflection response over the desired bandwidth are:

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0)} = \frac{200-50}{200+50} = 0.6\,, \quad |\Gamma_1|_{\max} = \frac{S_{\max}-1}{S_{\max}+1} = \frac{1.5-1}{1.5+1} = 0.2$$

It follows from Eq. (12.3.7) that the attenuation in dB over the desired band will be:

$$A = 20\log_{10}\left(\frac{|\Gamma_L|}{|\Gamma_1|_{\max}}\right) = 20\log_{10}\left(\frac{0.6}{0.2}\right) = 9.54 \text{ dB}$$

Because the number of sections and the attenuation are fixed, we may use the MATLAB function `chebtr3`. The following code segment calculates the various design parameters:

```
Z0 = 50; ZL = 200;
GL = z2g(ZL,Z0); Smax = 1.5;

f0 = 100; f = linspace(0,2*f0,401);          % plot over [0, 200] MHz

A = 20*log10(GL*(Smax+1)/(Smax-1));          % Eq. (12.3.8)

[Y,a,b,DF] = chebtr3(1/Z0, 1/ZL, 1, A);      % note, M = 1

Z = 1./Y; Df = f0*DF; L = 1/4;               % note, Z = [Z0, Z1, ZL]
```

```
G1 = abs(multiline(Z(1:2), L, ZL, f/f0));      % reflection response |Γ1(f)|

S = swr(G1);                                   % calculate SWR versus frequency

plot(f,G1); figure; plot(f,S);
```

The reflection response $|\Gamma_1(f)|$ is computed by `multiline` with frequencies normalized to the desired operating frequency of $f_0 = 100$ MHz. The impedance inputs to `multiline` were $[Z_0, Z_1]$ and Z_L and the electrical length of the segment was $L = 1/4$. The resulting bandwidth is $\Delta f = 35.1$ MHz. The reflection polynomials are:

$$\mathbf{b} = [b_0, b_1] = [\rho_1, \rho_1]\,, \quad \mathbf{a} = [a_0, a_1] = [1, \rho_1^2]\,, \quad \rho_1 = \frac{Z_1 - Z_0}{Z_1 + Z_0} = \frac{1}{3}$$

Two alternative ways to compute the reflection response are by using MATLAB's built-in function `freqz`, or the function `dtft`:

```
delta = pi * f/f0/2;
G1 = abs(freqz(b,a,2*delta));
% G1 = abs(dtft(b,2*delta) ./ dtft(a,2*delta));
```

where $2\delta = \pi f/f_0$ is the digital frequency, such that $z = e^{2j\delta}$. The bandwidth Δf can be computed from Eqs. (12.3.4) and (12.3.5), that is,

$$A = 10\log_{10}\left(\frac{x_0^2 + e_0^2}{1 + e_0^2}\right) \quad\Rightarrow\quad x_0 = \sqrt{(1+e_0^2)\,10^{A/10} - e_0^2}\,, \quad \Delta f = f_0\,\frac{4}{\pi}\,\mathrm{asin}\left(\frac{1}{x_0}\right)$$

where we replaced $T_1(x_0) = x_0$. □

Example 12.3.2: *Three- and four-section quarter-wavelength Chebyshev transformers.* Design a Chebyshev transformer that will match a 200-ohm load to a 50-ohm line. The line SWR is required to remain less than 1.25 over the frequency band $[50, 150]$ MHz.

Repeat the design if the SWR is required to remain less than 1.1 over the same bandwidth.

Solution: Here, we let the design specifications determine the number of sections and their characteristic impedances. In both cases, the unmatched reflection coefficient is the same as in the previous example, $\Gamma_L = 0.6$. Using $S_{\max} = 1.25$, the required attenuation in dB is for the first case:

$$A = 20\log_{10}\left(|\Gamma_L|\,\frac{S_{\max}+1}{S_{\max}-1}\right) = 20\log_{10}\left(0.6\,\frac{1.25+1}{1.25-1}\right) = 14.65\text{ dB}$$

The reflection coefficient corresponding to $S_{\max}$ is $|\Gamma_1|_{\max} = (1.25-1)/(1.25+1) = 1/9 = 0.1111$. In the second case, we use $S_{\max} = 1.1$ to find $A = 22.0074$ dB and $|\Gamma_1|_{\max} = (1.1-1)/(1.1+1) = 1/21 = 0.0476$.

In both cases, the operating frequency is at the middle of the given bandwidth, that is, $f_0 = 100$ MHz. The normalized bandwidth is $\Delta F = \Delta f/f_0 = (150-50)/100 = 1$. With these values of $A, \Delta F$, the function `chebtr` calculates the required number of sections and their impedances. The typical code is as follows:

```
Z0 = 50; ZL = 200;
GL = z2g(ZL,Z0); Smax = 1.25;

f1 = 50; f2 = 150;                               % given bandedge frequencies
Df = f2-f1; f0 = (f2+f1)/2; DF = Df/f0;          % operating frequency and bandwidth

A  = 20*log10(GL*(Smax+1)/(Smax-1));             % attenuation of reflectionless band

[Y,a,b] = chebtr(1/Z0, 1/ZL, A, DF);             % Chebyshev transformer design

Z = 1./Y; rho = n2r(Y);                          % impedances and reflection coefficients
```

For the first case, the resulting number of sections is $M = 3$, and the corresponding output vector of impedances Z, reflection coefficients at the interfaces, and reflection polynomials $\mathbf{a}, \mathbf{b}$ are:

$$\begin{aligned}
\boldsymbol{Z} &= [Z_0, Z_1, Z_2, Z_3, Z_L] = [50,\ 66.4185,\ 100,\ 150.5604,\ 200] \\
\boldsymbol{\rho} &= [\rho_1, \rho_2, \rho_3, \rho_4] = [0.1410,\ 0.2018,\ 0.2018,\ 0.1410] \\
\mathbf{b} &= [b_0, b_1, b_2, b_3] = [0.1410,\ 0.2115,\ 0.2115,\ 0.1410] \\
\mathbf{a} &= [a_0, a_1, a_2, a_3] = [1,\ 0.0976,\ 0.0577,\ 0.0199]
\end{aligned}$$

In the second case, we find $M = 4$ sections with design parameters:

$$\begin{aligned}
\boldsymbol{Z} &= [Z_0, Z_1, Z_2, Z_3, Z_4, Z_L] = [50,\ 59.1294,\ 81.7978,\ 122.2527,\ 169.1206,\ 200] \\
\boldsymbol{\rho} &= [\rho_1, \rho_2, \rho_3, \rho_4, \rho_5] = [0.0837,\ 0.1609,\ 0.1983,\ 0.1609,\ 0.0837] \\
\mathbf{b} &= [b_0, b_1, b_2, b_3, b_4] = [0.0837,\ 0.1673,\ 0.2091,\ 0.1673,\ 0.0837] \\
\mathbf{a} &= [a_0, a_1, a_2, a_3, a_4] = [1,\ 0.0907,\ 0.0601,\ 0.0274,\ 0.0070]
\end{aligned}$$

The reflection responses and SWRs are plotted versus frequency in Fig. 12.3.3. The upper two graphs corresponds to the case, $S_{\max} = 1.25$, and the bottom two graphs, to the case $S_{\max} = 1.1$.

The reflection responses $|\Gamma_1(f)|$ can be computed either with the help of the function `multiline`, or as the ratio of the reflection polynomials:

$$\Gamma_1(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{a_0 + a_1 z^{-1} + \cdots + a_M z^{-M}}\,, \quad z = e^{2j\delta}, \quad \delta = \frac{\pi}{2}\frac{f}{f_0}$$

The typical MATLAB code for producing these graphs uses the outputs of `chebtr`:

```
f = linspace(0,2*f0,401);                           % plot over [0, 200] MHz

M = length(Z)-2;                                    % number of sections
L = ones(1,M)/4;                                    % quarter-wave lengths
G1 = abs(multiline(Z(1:M+1), L, ZL, f/f0));         % ZL is a separate input

G1 = abs(freqz(b, a, pi*f/f0));                     % alternative way of computing G1

S = swr(G1);                                        % SWR on the line

plot(f,G1); figure; plot(f,S);
```

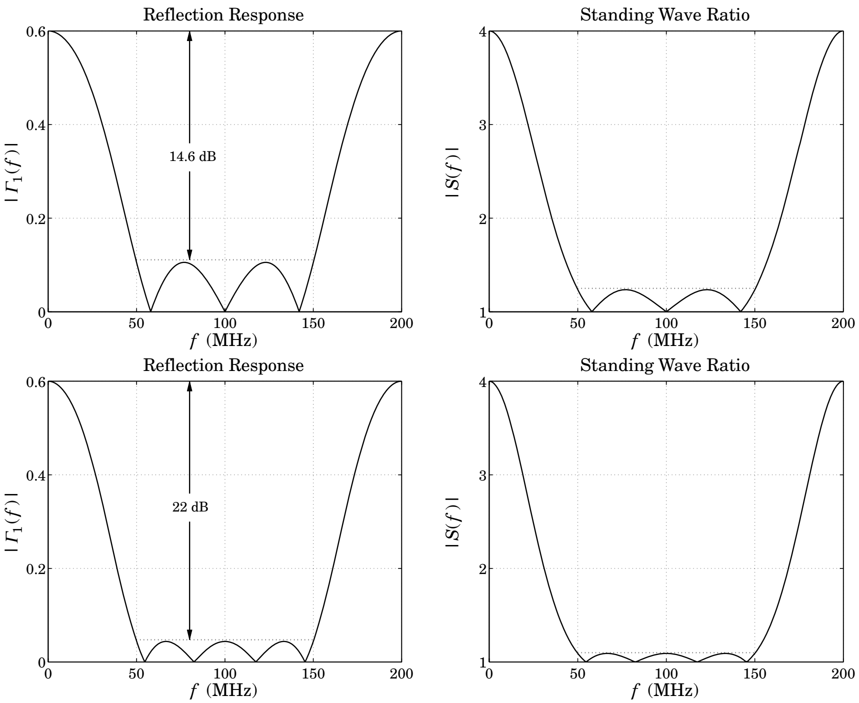


Fig. 12.3.3 Three and four section transformers.

In both cases, the section impedances satisfy the symmetry properties (12.3.9) and the reflection coefficients $\boldsymbol{\rho}$ are symmetric about their middle, as discussed in Sec. 6.8.

We note that the reflection coefficients ρ_i at the interfaces agree fairly closely with the reflection polynomial $\mathbf{b}$—equating the two is equivalent to the so-called *small-reflection approximation* that is usually made in designing quarter-wavelength transformers [781]. The above values are exact and do not depend on any approximation. □

12.4 Two-Section Dual-Band Chebyshev Transformers

Recently, a two-section sixth-wavelength transformer has been designed [952,953] that achieves matching at a frequency f_1 *and* its first harmonic $2f_1$. Each section has length $\lambda/6$ at the design frequency f_1. Such dual-band operation is desirable in certain applications, such as GSM and PCS systems. The transformer is depicted in Fig. 12.4.1.

Here, we point out that this design is actually equivalent to a two-section quarter-wavelength Chebyshev transformer whose parameters have been adjusted to achieve reflectionless notches at both frequencies f_1 and $2f_1$.

Using the results of the previous section, a two-section Chebyshev transformer will have reflection response:

$$|\Gamma_1(f)|^2 = \frac{e_1^2 T_2^2(x_0\cos\delta)}{1+e_1^2 T_2^2(x_0\cos\delta)}\,,\quad \delta = \frac{\pi}{2}\frac{f}{f_0} \tag{12.4.1}$$

where f_0 is the frequency at which the sections are quarter-wavelength. The second-order Chebyshev polynomial is $T_2(x)=2x^2-1$ and has roots at $x=\pm 1/\sqrt{2}$. We require that these two roots correspond to the frequencies f_1 and $2f_1$, that is, we set:

$$x_0\cos\delta_1 = \frac{1}{\sqrt{2}}\,,\quad x_0\cos 2\delta_1 = -\frac{1}{\sqrt{2}}\,,\quad \delta_1 = \frac{\pi}{2}\frac{f_1}{f_0} \tag{12.4.2}$$

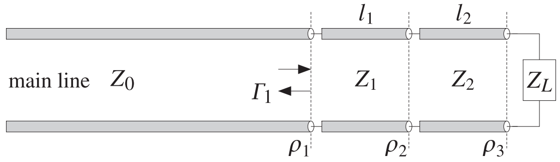


Fig. 12.4.1 Two-section dual-band Chebyshev transformer.

These conditions have the unique solution (such that $x_0 \geq 1$):

$$x_0 = \sqrt{2}\,,\quad \delta_1 = \frac{\pi}{3} = \frac{\pi}{2}\frac{f_1}{f_0} \quad\Rightarrow\quad f_0 = \frac{3}{2}f_1 \tag{12.4.3}$$

Thus, at f_1 the phase length is $\delta_1 = \pi/3 = 2\pi/6$, which corresponds to section lengths of $l_1 = l_2 = \lambda_1/6$, where $\lambda_1 = v/f_1$, and v is the propagation speed. Defining also $\lambda_0 = v/f_0$, we note that $\lambda_0 = 2\lambda_1/3$. According to Sec. 6.6, the most general two-section reflection response is expressed as the ratio of the second-order polynomials:

$$\Gamma_1(f) = \frac{B_1(z)}{A_1(z)} = \frac{\rho_1 + \rho_2(1+\rho_1\rho_3)z^{-1} + \rho_3 z^{-2}}{1+\rho_2(\rho_1+\rho_3)z^{-1}+\rho_1\rho_3 z^{-2}} \tag{12.4.4}$$

where

$$z = e^{2j\delta}\,,\quad \delta = \frac{\pi}{2}\frac{f}{f_0} = \frac{\pi}{3}\frac{f}{f_1} \tag{12.4.5}$$

and we used the relationship $2f_0 = 3f_1$ to express δ in terms of f_1. The polynomial $B_1(z)$ must have zeros at $z = e^{2j\delta_1} = e^{2\pi j/3}$ and $z = e^{2j(2\delta_1)} = e^{4\pi j/3} = e^{-2\pi j/3}$, hence, it must be (up to the factor ρ_1):

$$B_1(z) = \rho_1\big(1-e^{2\pi j/3}z^{-1}\big)\big(1-e^{-2\pi j/3}z^{-1}\big) = \rho_1(1+z^{-1}+z^{-2}) \tag{12.4.6}$$

Comparing this with (12.4.4), we arrive at the conditions:

$$\rho_3 = \rho_1\,,\quad \rho_2(1+\rho_1\rho_3) = \rho_1 \quad\Rightarrow\quad \rho_2 = \frac{\rho_1}{1+\rho_1^2} \tag{12.4.7}$$

We recall from the previous section that the condition $\rho_1 = \rho_3$ is equivalent to $Z_1Z_2 = Z_0Z_L$. Using (12.4.7) and the definition $\rho_2 = (Z_2 - Z_1)/(Z_2 + Z_1)$, or its inverse, $Z_2 = Z_1(1+\rho_2)/(1-\rho_2)$, we have:

$$Z_L Z_0 = Z_1 Z_2 = Z_1^2\,\frac{1+\rho_2}{1-\rho_2} = Z_1^2\,\frac{\rho_1^2+\rho_1+1}{\rho_1^2-\rho_1+1} = Z_1^2\,\frac{3Z_1^2+Z_0^2}{Z_1^2+3Z_0^2} \tag{12.4.8}$$

where in the last equation, we replaced $\rho_1 = (Z_1 - Z_0)/(Z_1 + Z_0)$. This gives a quadratic equation in Z_1^2. Picking the positive solution of the quadratic equation, we find:

$$Z_1 = \sqrt{\frac{Z_0}{6}\left[Z_L - Z_0 + \sqrt{(Z_L - Z_0)^2 + 36 Z_L Z_0}\right]} \tag{12.4.9}$$

Once Z_1 is known, we may compute $Z_2 = Z_L Z_0 / Z_1$. Eq. (12.4.9) is equivalent to the expression given by Monzon [953].

The sections are quarter-wavelength at f_0 and sixth-wavelength at f_1, that is, $l_1 = l_2 = \lambda_1/6 = \lambda_0/4$. We note that the frequency f_0 lies exactly in the middle between f_1 and $2f_1$. Viewed as a quarter-wavelength transformer, the bandwidth will be:

$$\sin\left(\frac{\pi}{4}\frac{\Delta f}{f_0}\right) = \frac{1}{x_0} = \frac{1}{\sqrt{2}} \quad\Rightarrow\quad \Delta f = f_0 = 1.5 f_1 \tag{12.4.10}$$

which spans the interval $[f_0 - \Delta f/2, f_0 + \Delta f/2] = [0.75 f_1, 2.25 f_1]$. Using $T_2(x_0) = 2x_0^2 - 1 = 3$ and Eq. (12.3.6), we find the attenuation achieved over the bandwidth Δf:

$$\sqrt{(1 + e_0^2)10^{A/10} - e_0^2} = T_2(x_0) = 3 \quad\Rightarrow\quad A = 10\log_{10}\left(\frac{9 + e_0^2}{1 + e_0^2}\right) \tag{12.4.11}$$

As an example, we consider the matching of $Z_L = 200\ \Omega$ to $Z_0 = 50\ \Omega$. The section impedances are found from Eq. (12.4.9) to be: $Z_1 = 80.02\ \Omega$, $Z_2 = 124.96\ \Omega$. More simply, we can invoke the function `chebtr2` with $M = 2$ and $\Delta F = \Delta f / f_0 = 1$.

Fig. 12.4.2 shows the designed reflection response normalized to its dc value, that is, $|\Gamma_1(f)|^2/|\Gamma_1(0)|^2$. The response has exact zeros at f_1 and $2f_1$. The attenuation was $A = 7.9$ dB. The reflection coefficients were $\rho_1 = \rho_3 = 0.2309$ and $\rho_2 = \rho_1/(1 + \rho_1^2) = 0.2192$, and the reflection polynomials:

$$B_1(z) = 0.2309(1 + z^{-1} + z^{-2}), \quad A_1(z) = 1 + 0.1012 z^{-1} + 0.0533 z^{-2}$$

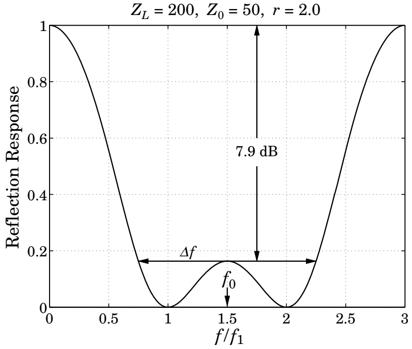


Fig. 12.4.2 Reflection response $|\Gamma_1(f)|^2$ normalized to unity gain at dc.

The reflection response can be computed using Eq. (12.4.1), or using the MATLAB function `multiline`, or the function `freqz` and the computed polynomial coefficients. The following code illustrates the computation using `chebtr2`:

```
Z0 = 50; ZL = 100; x0 = sqrt(2); e0sq = (ZL-Z0)^2/(4*ZL*Z0); e1sq = e0sq/9;

[Y,a1,b1,A] = chebtr2(1/Z0, 1/ZL, 2, 1);                 % a1  =  [1, 0.1012, 0.0533]
                                                         % b1  =  [0.2309, 0.2309, 0.2309]
Z = 1./Y; rho = n2r(Z0*Y);                               % Z  =  [50, 80.02, 124.96, 200]
                                                         % ρ  =  [0.2309, 0.2192, 0.2309]
f = linspace(0,3,301);                                   % f is in units of f1
delta = pi*f/3;  x = x0*cos(delta); T2 = 2*x.^2-1;

G1 = e1sq*T2.^2 ./ (1 + e1sq*T2.^2);

% G1 = abs(multiline(Z(1:3), [1,1]/6, ZL, f)).^2;        % alternative calculation
% G1 = abs(freqz(b1,a1, 2*delta)).^2;                    % alternative calculation
% G1 = abs(dtft(b1,2*delta)./dtft(a1,2*delta)).^2;       % alternative calculation

plot(f, G1/G1(1));
```

The above design method is not restricted to the first and second harmonics. It can be generalized to any two frequencies f_1, f_2 at which the two-section transformer is required to be reflectionless [954,955].

Possible applications are the matching of dual-band antennas operating in the cellular/PCS, GSM/DCS, WLAN, GPS, and ISM bands, and other dual-band RF applications for which the frequency f_2 is not necessarily $2f_1$.

We assume that $f_1 < f_2$, and define $r = f_2/f_1$, where r can take any value greater than unity. The reflection polynomial $B_1(z)$ is constructed to have zeros at f_1, f_2:

$$B_1(z) = \rho_1(1 - e^{2j\delta_1}z^{-1})(1 - e^{2j\delta_2}z^{-1}), \quad \delta_1 = \frac{\pi f_1}{2f_0}, \ \delta_2 = \frac{\pi f_2}{2f_0} \tag{12.4.12}$$

The requirement that the segment impedances, and hence the reflection coefficients ρ_1, ρ_2, ρ_3, be real-valued implies that the zeros of $B_1(z)$ must be conjugate pairs. This can be achieved by choosing the quarter-wavelength normalization frequency f_0 to lie half-way between f_1, f_2, that is, $f_0 = (f_1 + f_2)/2 = (r + 1)f_1/2$. This implies that:

$$\boxed{\delta_1 = \frac{\pi}{r+1}, \quad \delta_2 = r\delta_1 = \pi - \delta_1} \tag{12.4.13}$$

The phase length at any frequency f will be:

$$\delta = \frac{\pi}{2}\frac{f}{f_0} = \frac{\pi}{r+1}\frac{f}{f_1} \tag{12.4.14}$$

The section lengths become quarter-wavelength at f_0 and $2(r+1)$-th wavelength at f_1:

$$l_1 = l_2 = \frac{\lambda_0}{4} = \frac{\lambda_1}{2(r+1)} \tag{12.4.15}$$

It follows now from Eq. (12.4.13) that the zeros of $B_1(z)$ are complex-conjugate pairs:

$$e^{2j\delta_2} = e^{2j(\pi - \delta_1)} = e^{-2j\delta_1} \tag{12.4.16}$$

Then, $B_1(z)$ takes the form:

$$B_1(z) = \rho_1(1 - e^{2j\delta_1}z^{-1})(1 - e^{-2j\delta_1}z^{-1}) = \rho_1(1 - 2\cos 2\delta_1\, z^{-1} + z^{-2}) \tag{12.4.17}$$

Comparing with Eq. (12.4.4), we obtain the reflection coefficients:

$$\rho_3 = \rho_1\,, \quad \rho_2 = -\frac{2\rho_1\cos 2\delta_1}{1+\rho_1^2} \tag{12.4.18}$$

Proceeding as in (12.4.8) and using the identity $\tan^2\delta_1 = (1-\cos 2\delta_1)/(1+\cos 2\delta_1)$, we find the following equation for the impedance Z_1 of the first section:

$$Z_L Z_0 = Z_1 Z_2 = Z_1^2\,\frac{1+\rho_2}{1-\rho_2} = Z_1^2\,\frac{\rho_1^2 - 2\rho_1\cos 2\delta_1 + 1}{\rho_1^2 + 2\rho_1\cos 2\delta_1 + 1} = Z_1^2\,\frac{Z_1^2\tan^2\delta_1 + Z_0^2}{Z_1^2 + Z_0^2\tan^2\delta_1} \tag{12.4.19}$$

with solution for Z_1 and Z_2:

$$Z_1 = \sqrt{\frac{Z_0}{2\tan^2\delta_1}\left[Z_L - Z_0 + \sqrt{(Z_L-Z_0)^2 + 4Z_L Z_0\tan^4\delta_1}\,\right]}\,, \quad Z_2 = \frac{Z_0 Z_L}{Z_1} \tag{12.4.20}$$

Equations (12.4.13), (12.4.15), and (12.4.20) provide a complete solution to the two-section transformer design problem. The design equations have been implemented by the MATLAB function `dualband`:

```
[Z1,Z2,a1,b1] = dualband(Z0,ZL,r);        % two-section dual-band Chebyshev transformer
```

where $\mathbf{a}_1, \mathbf{b}_1$ are the coefficients of $A_1(z)$ and $B_1(z)$. Next, we show that $B_1(z)$ is indeed proportional to the Chebyshev polynomial $T_2(x)$. Setting $z = e^{2j\delta}$, where δ is given by (12.4.14), we find:

$$\begin{aligned} B_1(z) &= \rho_1(z + z^{-1} - 2\cos 2\delta_1)z^{-1} = \rho_1(2\cos 2\delta - 2\cos 2\delta_1)e^{-2j\delta} \\ &= 4\rho_1(\cos^2\delta - \cos^2\delta_1)e^{-2j\delta} = 4\rho_1\cos^2\delta_1\big(\frac{\cos^2\delta}{\cos^2\delta_1} - 1\big)e^{-2j\delta} \\ &= 4\rho_1\cos^2\delta_1(2x_0^2\cos^2\delta - 1)e^{-2j\delta} = 4\rho_1\cos^2\delta_1 T_2(x_0\cos\delta)e^{-2j\delta} \end{aligned} \tag{12.4.21}$$

where we defined:

$$x_0 = \frac{1}{\sqrt{2}\cos\delta_1} \tag{12.4.22}$$

We may also show that the reflection response $|\Gamma_1(f)|^2$ is given by Eq. (12.4.1). At zero frequency, $\delta = 0$, we have $T_2(x_0) = 2x_0^2 - 1 = \tan^2\delta_1$. As discussed in Sec. 6.8, the sum of the coefficients of the polynomial $B_1(z)$, or equivalently, its value at dc, $\delta = 0$ or $z = 1$, must be given by $|B_1(1)|^2 = \sigma^2 e_0^2$, where

$$\sigma^2 = (1-\rho_1^2)(1-\rho_2^2)(1-\rho_3^2)\,, \quad e_0^2 = \frac{(Z_L - Z_0)^2}{4Z_L Z_0} \tag{12.4.23}$$

Using Eq. (12.4.21), this condition reads $\sigma^2 e_0^2 = |B_1(1)|^2 = 16\rho_1^2\cos^4\delta_1 T_2^2(x_0)$, or, $\sigma^2 e_0^2 = 16\rho_1^2\sin^4\delta_1$. This can be verified with some tedious algebra. Because $e_1^2 = e_0^2/T_2^2(x_0)$, the same condition reads $\sigma^2 e_1^2 = 16\rho_1^2\cos^4\delta_1$.

It follows that $|B_1(z)|^2 = \sigma^2 e_1^2 T_2^2(x)$. On the other hand, according to Sec. 6.6, the denominator polynomial $A_1(z)$ in (12.4.4) satisfies $|A_1(z)|^2 - |B_1(z)|^2 = \sigma^2$, or, $|A_1(z)|^2 = \sigma^2 + |B_1(z)|^2$. Therefore,

$$|\Gamma_1(f)|^2 = \frac{|B_1(z)|^2}{|A_1(z)|^2} = \frac{|B_1(z)|^2}{\sigma^2 + |B_1(z)|^2} = \frac{\sigma^2 e_1^2 T_2^2(x)}{\sigma^2 + \sigma^2 e_1^2 T_2^2(x)} = \frac{e_1^2 T_2^2(x)}{1 + e_1^2 T_2^2(x)} \tag{12.4.24}$$

Thus, the reflectance is identical to that of a two-section Chebyshev transformer. However, the interpretation as a quarter-wavelength transformer, that is, a transformer whose attenuation at f_0 is less than the attenuation at dc, is valid only for a limited range of values, that is, $1 \le r \le 3$. For this range, the parameter x_0 defined in (12.4.22) is $x_0 \ge 1$. In this case, the corresponding bandwidth about f_0 can be meaningfully defined through Eq. (12.3.4), which gives:

$$\sin\left(\frac{\pi}{2(r+1)}\,\frac{\Delta f}{f_1}\right) = \sqrt{2}\cos\delta_1 = \sqrt{2}\cos\left(\frac{\pi}{r+1}\right) \tag{12.4.25}$$

For $1 \le r \le 3$, the right-hand side is always less than unity. On the other hand, when $r > 3$, the parameter x_0 becomes $x_0 < 1$, the bandwidth Δf loses its meaning, and the reflectance at f_0 becomes greater than that at dc, that is, a gain. For any value of r, the attenuation or gain at f_0 can be calculated from Eq. (12.3.5) with $M = 2$:

$$A = 10\log_{10}\left(\frac{T_2^2(x_0) + e_0^2}{1 + e_0^2}\right) = 10\log_{10}\left(\frac{\tan^4\delta_1 + e_0^2}{1 + e_0^2}\right) \tag{12.4.26}$$

The quantity A is positive for $1 < r < 3$ or $\tan\delta_1 > 1$, and negative for $r > 3$ or $\tan\delta_1 < 1$. For the special case of $r = 3$, we have $\delta_1 = \pi/4$ and $\tan\delta_1 = 1$, which gives $A = 0$. Also, it follows from (12.4.18) that $\rho_2 = 0$, which means that $Z_1 = Z_2$ and (12.4.19) gives $Z_1^2 = Z_L Z_0$. The two sections combine into a single section of double length $2l_1 = \lambda_1/4$ at f_1, that is, a single-section quarter wavelength transformer, which, as is well known, has zeros at odd multiples of its fundamental frequency.

For the case $r = 2$, we have $\delta_1 = \pi/3$ and $\tan\delta_1 = \sqrt{3}$. The design equation (12.4.20) reduces to that given in [953] and the section lengths become $\lambda_1/6$.

Fig. 12.4.3 shows two examples, one with $r = 2.5$ and one with $r = 3.5$, both transforming $Z_L = 200$ into $Z_0 = 50$ ohm.

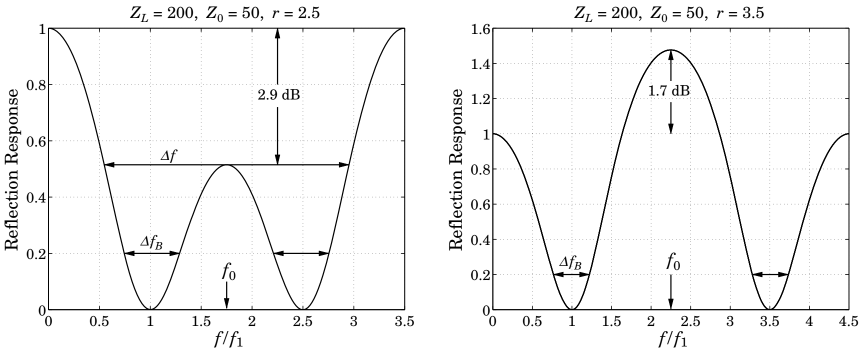


Fig. 12.4.3 Dual-band transformers at frequencies $\{f_1, 2.5f_1\}$ and $\{f_1, 3.5f_1\}$.

The reflectances are normalized to unity gain at dc. For $r = 2.5$, we find $Z_1 = 89.02$ and $Z_2 = 112.33$ ohm, and attenuation $A = 2.9$ dB. The section lengths at f_1 are $l_1 = l_2 = \lambda_1/(2(2.5+1)) = \lambda_1/7$. The bandwidth Δf calculated from Eq. (12.4.25) is shown on the left graph. For the case $r = 3.5$, we find $Z_1 = 112.39$ and $Z_2 = 88.98$ ohm and section lengths $l_1 = l_2 = \lambda_1/9$. The quantity A is negative, $A = -1.7$ dB, signifying a gain at f_0. The polynomial coefficients were in the two cases:

$$\begin{aligned} r = 2.5,\quad \mathbf{a}_1 &= [1,\ \ 0.0650,\ 0.0788], & \mathbf{b}_1 &= [0.2807,\ \ 0.1249,\ 0.2807] \\ r = 3.5,\quad \mathbf{a}_1 &= [1,\ -0.0893,\ 0.1476], & \mathbf{b}_1 &= [0.3842,\ -0.1334,\ 0.3842] \end{aligned}$$

The bandwidth about f_1 and f_2 corresponding to any desired bandwidth level can be obtained in closed form. Let Γ_B be the desired bandwidth level. Equivalently, Γ_B can be determined from a desired SWR level S_B through $\Gamma_B = (S_B - 1)/(S_B + 1)$. The bandedge frequencies can be derived from Eq. (12.4.24) by setting:

$$|\Gamma_1(f)|^2 = \Gamma_B^2$$

Solving this equation, we obtain the left and right bandedge frequencies:

$$\begin{aligned} f_{1L} &= \frac{2f_0}{\pi}\,\mathrm{asin}\big(\sqrt{1-a}\,\sin\delta_1\big)\,, \qquad & f_{2R} &= 2f_0 - f_{1L} \\ f_{1R} &= \frac{2f_0}{\pi}\,\mathrm{asin}\big(\sqrt{1+a}\,\sin\delta_1\big)\,, \qquad & f_{2L} &= 2f_0 - f_{1R} \end{aligned} \tag{12.4.27}$$

where $f_0 = (f_1 + f_2)/2$ and a is defined in terms of Γ_B and Γ_L by:

$$a = \left[\frac{\Gamma_B^2}{1-\Gamma_B^2}\,\frac{1-\Gamma_L^2}{\Gamma_L^2}\right]^{1/2} = \frac{S_B - 1}{S_L - 1}\sqrt{\frac{S_L}{S_B}} \tag{12.4.28}$$

where $\Gamma_L = (Z_L - Z_0)/(Z_L + Z_0)$ and $S_L = (1 + |\Gamma_L|)/(1 - |\Gamma_L|)$. We note the symmetry relations: $f_{1L} + f_{2R} = f_{1R} + f_{2L} = 2f_0$. These imply that the bandwidths about f_1 and f_2 are the same:

$$\Delta f_B = f_{1R} - f_{1L} = f_{2R} - f_{2L} \tag{12.4.29}$$

The MATLAB function `dualbw` implements Eqs. (12.4.27):

```
[f1L,f1R,f2L,f2R] = dualbw(ZL,Z0,r,GB);        % bandwidths of dual-band transformer
```

The bandwidth Δf_B is shown in Fig. 12.4.3. For illustration purposes, it was computed at a level such that $\Gamma_B^2/\Gamma_L^2 = 0.2$.

12.5 Quarter-Wavelength Transformer With Series Section

One limitation of the Chebyshev quarter-wavelength transformer is that it requires the load to be real-valued. The method can be modified to handle complex loads, but generally the wide bandwidth property is lost. The modification is to insert the quarter-wavelength transformer not at the load, but at a distance from the load corresponding to a voltage minimum or maximum.

For example, Fig. 12.5.1 shows the case of a single quarter-wavelength section inserted at a distance $L_{\min}$ from the load. At that point, the wave impedance seen by the quarter-wave transformer will be real-valued and given by $Z_{\min} = Z_0/S_L$, where S_L is the SWR of the unmatched load. Alternatively, one can choose a point of voltage maximum $L_{\max}$ at which the wave impedance will be $Z_{\max} = Z_0 S_L$.

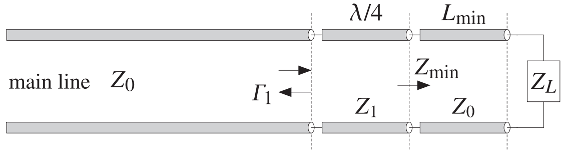


Fig. 12.5.1 Quarter-wavelength transformer for matching a complex load.

As we saw in Sec. 10.13, the electrical lengths $L_{\min}$ or $L_{\max}$ are related to the phase angle θ_L of the load reflection coefficient Γ_L by Eqs. (10.13.2) and (10.13.3). The MATLAB function `lmin` can be called to calculate these distances and corresponding wave impedances.

The calculation of the segment length, $L_{\min}$ or $L_{\max}$, depends on the desired matching frequency f_0. Because a complex impedance can vary rapidly with frequency, the segment will have the wrong length at other frequencies.

Even if the segment is followed by a multisection transformer, the presence of the segment will tend to restrict the overall operating bandwidth to essentially that of a single quarter-wavelength section. In the case of a single section, its impedance can be calculated simply as:

$$Z_1 = \sqrt{Z_0 Z_{\min}} = \frac{1}{\sqrt{S_L}}\,Z_0 \quad \text{and} \quad Z_1 = \sqrt{Z_0 Z_{\max}} = \sqrt{S_L}\,Z_0 \tag{12.5.1}$$

Example 12.5.1: *Quarter-wavelength matching of a complex load impedance.* Design a quarter-wavelength transformer of length $M = 1, 3, 5$ that will match the complex impedance $Z_L = 200 + j100$ ohm to a 50-ohm line at $f_0 = 100$ MHz. Perform the design assuming the maximum reflection coefficient level of $|\Gamma_1|_{\max} = 0.1$.

Assuming that the inductive part of Z_L arises from an inductance, replace the complex load by $Z_L = 200 + j100f/f_0$ at other frequencies. Plot the corresponding reflection response $|\Gamma_1(f)|$ versus frequency.

Solution: At f_0, the load is $Z_L = 200 + j100$ and its reflection coefficient and SWR are found to be $|\Gamma_L| = 0.6695$ and $S_L = 5.0521$. It follows that the line segments corresponding to a voltage minimum and maximum will have parameters:

$$L_{\min} = 0.2665, \quad Z_{\min} = \frac{1}{S_L} Z_0 = 9.897, \qquad L_{\max} = 0.0165, \quad Z_{\max} = S_L Z_0 = 252.603$$

For either of these cases, the effective load reflection coefficient seen by the transformer will be $|\Gamma| = (S_L - 1)/(S_L + 1) = 0.6695$. It follows that the design attenuation specification for the transformer will be:

$$A = 20\log_{10}\left(\frac{|\Gamma|}{|\Gamma_1|_{\max}}\right) = 20\log_{10}\left(\frac{0.6695}{0.1}\right) = 16.5155 \text{ dB}$$

With the given number of sections M and this value of the attenuation A, the following MATLAB code will design the transformer and calculate the reflection response of the overall structure:

```
Z0 = 50; ZL0 = 200 + 100j;                        % load impedance at f0

[Lmin, Zmin] = lmin(ZL0,Z0,'min');                % calculate Lmin

Gmin = abs(z2g(Zmin,Z0)); G1max = 0.1;            % design based on Zmin
A = 20*log10(Gmin/G1max);

M = 3;                                            % three-section transformer
Z = 1./chebtr3(1/Z0, 1/Zmin, M, A);
Ztot = [Z(1:M+1), Z0];                            % concatenate all sections
Ltot = [ones(1,M)/4, Lmin];                       % electrical lengths of all sections

f0 = 100; f = linspace(0,2*f0, 801);
ZL = 200 + j*100*f/f0;                            % assume inductive load

G1 = abs(multiline(Ztot, Ltot, ZL, f/f0));        % overall reflection response
```

where the designed impedances and quarter-wavelength segments are concatenated with the last segment of impedance Z_0 and length $L_{\min}$ or $L_{\max}$. The corresponding frequency reflection responses are shown in Fig. 12.5.2.

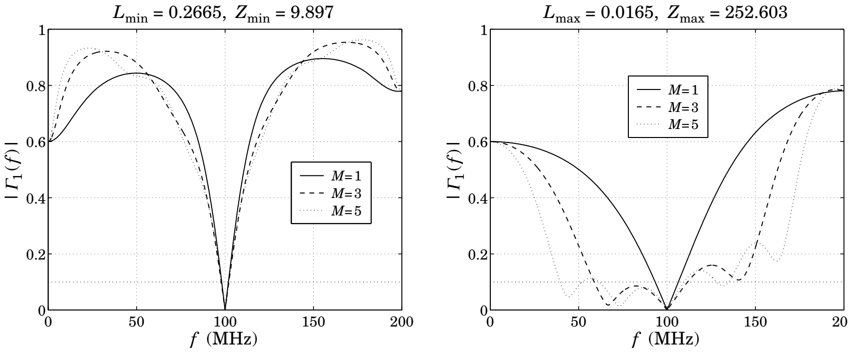


Fig. 12.5.2 Matching a complex impedance.

The calculated vector outputs of the transformer impedances are in the $L_{\min}$ case:

$$\mathbf{Z} = [50,\ 50/S_L^{1/2},\ 50/S_L] = [50,\ 22.2452,\ 9.897]$$

$$\mathbf{Z} = [50,\ 36.5577,\ 22.2452,\ 13.5361,\ 9.897]$$

$$\mathbf{Z} = [50,\ 40.5325,\ 31.0371,\ 22.2452,\ 15.9437,\ 12.2087,\ 9.897]$$

and in the $L_{\max}$ case:

$$\mathbf{Z} = [50,\ 50S_L^{1/2},\ 50S_L] = [50,\ 112.3840,\ 252.603]$$

$$\mathbf{Z} = [50,\ 68.3850,\ 112.3840,\ 184.6919,\ 252.603]$$

$$\mathbf{Z} = [50,\ 61.6789,\ 80.5486,\ 112.3840,\ 156.8015,\ 204.7727,\ 252.603]$$

We note that there is essentially no difference in bandwidth over the desired design level of $|\Gamma_1|_{\max} = 0.1$ in the $L_{\min}$ case, and very little difference in the $L_{\max}$ case. □

The MATLAB function `qwt1` implements this matching method. Its inputs are the complex load and line impedances Z_L, Z_0 and its outputs are the quarter-wavelength section impedance Z_1 and the electrical length L_m of the Z_0-section. It has usage:

```
[Z1,Lm] = qwt1(ZL,Z0,type);        % λ/4-transformer with series section
```

where `type` is one of the strings `'min'` or `'max'`, depending on whether the first section gives a voltage minimum or maximum.

12.6 Quarter-Wavelength Transformer With Shunt Stub

Two other possible methods of matching a complex load are to use a shorted or opened stub connected in parallel with the load and adjusting its length or its line impedance so that its susceptance cancels the load susceptance, resulting in a real load that can then be matched by the quarter-wave section.

In the first method, the stub length is chosen to be either $\lambda/8$ or $3\lambda/8$ and its impedance is determined in order to provide the required cancellation of susceptance.

In the second method, the stub's characteristic impedance is chosen to have a convenient value and its length is determined in order to provide the susceptance cancellation.

These methods are shown in Fig. 12.6.1. In practice, they are mostly used with microstrip lines that have easily adjustable impedances. The methods are similar to the stub matching methods discussed in Sec. 12.8 in which the stub is not connected at the load but rather after the series segment.

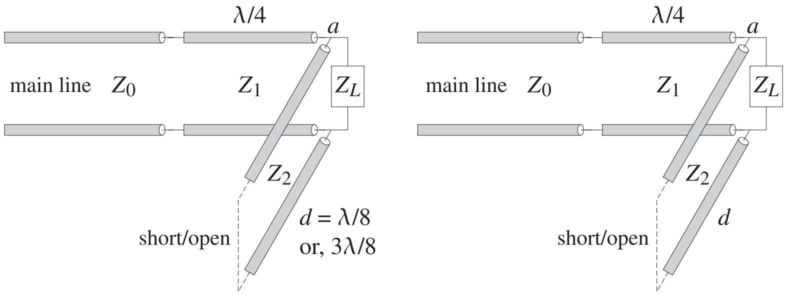


Fig. 12.6.1 Matching with a quarter-wavelength section and a shunt stub.

Let $Y_L = 1/Z_L = G_L + jB_L$ be the load admittance. The admittance of a shorted stub of characteristic admittance $Y_2 = 1/Z_2$ and length d is $Y_{\text{stub}} = -jY_2\cot\beta d$ and that of an opened stub, $Y_{\text{stub}} = jY_2\tan\beta d$.

The total admittance at point a in Fig. 12.6.1 is required to be real-valued, resulting in the susceptance cancellation condition:

$$Y_a = Y_L + Y_{\text{stub}} = G_L + j(B_L - Y_2\cot\beta d) = G_L \quad \Rightarrow \quad Y_2\cot\beta d = B_L \tag{12.6.1}$$

For an opened stub the condition becomes $Y_2\tan\beta d = -B_L$. In the first method, the stub length is $d = \lambda/8$ or $3\lambda/8$ with phase thicknesses $\beta d = \pi/4$ or $3\pi/4$. The corresponding values of the cotangents and tangents are $\cot\beta d = \tan\beta d = 1$ or $\cot\beta d = \tan\beta d = -1$.

Then, the susceptance cancellation condition becomes $Y_2 = B_L$ for a shorted $\lambda/8$-stub or an opened $3\lambda/8$-stub, and $Y_2 = -B_L$ for a shorted $3\lambda/8$-stub or an opened $\lambda/8$-stub. The case $Y_2 = B_L$ must be chosen when $B_L > 0$ and $Y_2 = -B_L$, when $B_L < 0$.

In the second method, Z_2 is chosen and the length d is determined from the condition (12.6.1), $\cot\beta d = B_L/Y_2 = Z_2 B_L$ for a shorted stub, and $\tan\beta d = -Z_2 B_L$ for an opened one. The resulting d must be reduced modulo $\lambda/2$ to a positive value.

With the cancellation of the load susceptance, the impedance looking to the right of point a will be real-valued, $Z_a = 1/Y_a = 1/G_L$. Therefore, the quarter-wavelength section will have impedance:

$$Z_1 = \sqrt{Z_0 Z_a} = \sqrt{\frac{Z_0}{G_L}} \tag{12.6.2}$$

The MATLAB functions `qwt2` and `qwt3` implement the two matching methods. Their usage is as follows:

```
[Z1,Z2] = qwt2(ZL,Z0);          % λ/4-transformer with λ/8 shunt stub
[Z1,d]  = qwt3(ZL,Z0,Z2,type)   % λ/4-transformer with shunt stub of given impedance
```

where `type` takes on the string values `'s'` or `'o'` for shorted or opened stubs.

Example 12.6.1: Design quarter-wavelength matching circuits to match the load impedance $Z_L = 15 + 20j\ \Omega$ to a 50-ohm generator at 5 GHz using series sections and shunt stubs. Use microstrip circuits with a Duroid substrate ($\epsilon_r = 2.2$) of height $h = 1$ mm. Determine the lengths and widths of all required microstrip sections, choosing always the shortest possible lengths.

Solution: For the quarter-wavelength transformer with a series section, it turns out that the shortest length corresponds to a voltage maximum. The impedance Z_1 and section length $L_{\max}$ are computed with the MATLAB function `qwt1`:

$$[Z_1, L_{\max}] = \text{qwt1}(Z_L, Z_0, \texttt{'max'}) \quad\Rightarrow\quad Z_1 = 98.8809\ \Omega, \quad L_{\max} = 0.1849$$

The widths and lengths of the microstrip sections are designed with the help of the functions `mstripr` and `mstripa`. For the quarter-wavelength section Z_1, the corresponding width-to-height ratio $u_1 = w_1/h$ is calculated from `mstripr` and then used in `mstripa` to get the effective permittivity, from which the wavelength and length of the segment can be calculated:

$$\begin{aligned}
u_1 &= \text{mstripr}(\epsilon_r, Z_1) = 0.9164, \quad w_1 = u_1 h = 0.9164\text{ mm}\\
\epsilon_{\text{eff}} &= \text{mstripa}(\epsilon_r, u_1) = 1.7659, \quad \lambda_1 = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 4.5151\text{ cm}, \quad l_1 = \frac{\lambda_1}{4} = 1.1288\text{ cm}
\end{aligned}$$

where the free-space wavelength is $\lambda_0 = 6$ cm. Similarly, we find for the series segment with impedance $Z_2 = Z_0$ and length $L_2 = L_{\max}$:

$$\begin{aligned}
u_2 &= \text{mstripr}(\epsilon_r, Z_2) = 3.0829, \quad w_2 = u_2 h = 3.0829\text{ mm}\\
\epsilon_{\text{eff}} &= \text{mstripa}(\epsilon_r, u_2) = 1.8813, \quad \lambda_2 = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 4.3745\text{ cm}, \quad l_2 = L_2\lambda_2 = 0.8090\text{ cm}
\end{aligned}$$

For the case of the $\lambda/8$ shunt stub, we find from `qwt2`:

$$[Z_1, Z_2] = \text{qwt2}(Z_L, Z_0) = [45.6435, -31.2500]\ \Omega$$

where the negative Z_2 means that we should use either a shorted $3\lambda/8$ stub or an opened $\lambda/8$ one. Choosing the latter and setting $Z_2 = 31.25\ \Omega$, we can go on to calculate the microstrip widths and lengths:

$$\begin{aligned}
u_1 &= \text{mstripr}(\epsilon_r, Z_1) = 3.5241, \quad w_1 = u_1 h = 3.5241\text{ mm}\\
\epsilon_{\text{eff}} &= \text{mstripa}(\epsilon_r, u_1) = 1.8965, \quad \lambda_1 = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 4.3569\text{ cm}, \quad l_1 = \frac{\lambda_1}{4} = 1.0892\text{ cm}\\
u_2 &= \text{mstripr}(\epsilon_r, Z_2) = 5.9067, \quad w_2 = u_2 h = 5.9067\text{ mm}\\
\epsilon_{\text{eff}} &= \text{mstripa}(\epsilon_r, u_2) = 1.9567, \quad \lambda_2 = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 4.2894\text{ cm}, \quad l_2 = \frac{\lambda_2}{8} = 0.5362\text{ cm}
\end{aligned}$$

For the third matching method, we use a shunt stub of impedance $Z_2 = 30\ \Omega$. It turns out that the short-circuited version has the shorter length. We find with the help of `qwt3`:

$$[Z_1, d] = \text{qwt3}(Z_L, Z_0, Z_2, \texttt{'s'}) \quad\Rightarrow\quad Z_1 = 45.6435\ \Omega, \quad d = 0.3718$$

The microstrip width and length of the quarter-wavelength section Z_1 are the same as in the previous case, because the two cases differ only in the way the load susceptance is canceled. The microstrip parameters of the shunt stub are:

$$\begin{aligned}
u_2 &= \text{mstripr}(\epsilon_r, Z_2) = 6.2258, \quad w_2 = u_2 h = 6.2258\text{ mm}\\
\epsilon_{\text{eff}} &= \text{mstripa}(\epsilon_r, u_2) = 1.9628, \quad \lambda_2 = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 4.2826\text{ cm}, \quad l_2 = d\lambda_2 = 1.5921\text{ cm}
\end{aligned}$$

Had we used a 50 Ω shunt segment, its width and length would be $w_2 = 3.0829$ mm and $l_2 = 1.7983$ cm. Fig. 12.6.2 depicts the microstrip matching circuits. □

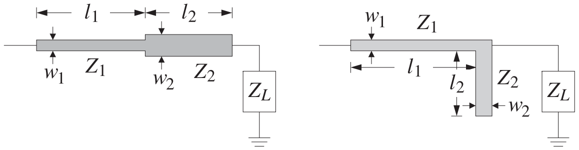


Fig. 12.6.2 Microstrip matching circuits.

12.7 *Two-Section Series Impedance Transformer*

One disadvantage of the quarter-wavelength transformer is that the required impedances of the line segments are not always easily realized. In certain applications, such as microwave integrated circuits, the segments are realized by microstrip lines whose impedances can be adjusted easily by changing the strip widths. In other applications, however, such as matching antennas to transmitters, we typically use standard 50- and 75-ohm coaxial cables and it is not possible to re-adjust their impedances.

The two-section series impedance transformer, shown in Fig. 12.7.1, addresses this problem [942,943]. It employs two line segments of *known* impedances Z_1 and Z_2 that have convenient values and adjusts their (electrical) lengths L_1 and L_2 to match a complex load Z_L to a main line of impedance Z_0. Fig. 12.7.1 depicts this kind of transformer.

The design method is identical to that of designing two-layer antireflection coatings discussed in Sec. 6.2. Here, we modify that method slightly in order to handle complex load impedances. We assume that Z_0, Z_1, and Z_2 are real and the load complex, $Z_L = R_L + jX_L$.

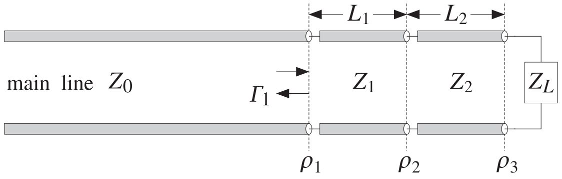


Fig. 12.7.1 Two-section series impedance transformer.

Defining the phase thicknesses of the two segments by $\delta_1 = 2\pi n_1 l_1/\lambda_0 = 2\pi L_1$ and $\delta_2 = 2\pi n_2 l_2/\lambda_0 = 2\pi L_2$, the reflection responses Γ_1 and Γ_2 at interfaces 1 and 2 are:

$$\Gamma_1 = \frac{\rho_1 + \Gamma_2 e^{-2j\delta_1}}{1+\rho_1\Gamma_2 e^{-2j\delta_1}}\,, \quad \Gamma_2 = \frac{\rho_2 + \rho_3 e^{-2j\delta_2}}{1+\rho_2\rho_3 e^{-2j\delta_2}}$$

where the elementary reflection coefficients are:

$$\rho_1 = \frac{Z_1 - Z_0}{Z_1 + Z_0}\,, \quad \rho_2 = \frac{Z_2 - Z_1}{Z_2 + Z_1}\,, \quad \rho_3 = \frac{Z_L - Z_2}{Z_L + Z_2}$$

The coefficients ρ_1, ρ_2 are real, but ρ_3 is complex, and we may represent it in polar form $\rho_3 = |\rho_3|e^{j\theta_3}$. The reflectionless matching condition is $\Gamma_1 = 0$ (at the operating free-space wavelength λ_0). This requires that $\rho_1 + \Gamma_2 e^{-2j\delta_1} = 0$, which implies:

$$e^{2j\delta_1} = -\frac{\Gamma_2}{\rho_1} \tag{12.7.1}$$

Because the left-hand side has unit magnitude, we must have the condition $|\Gamma_2| = |\rho_1|$, or, $|\Gamma_2|^2 = \rho_1^2$, which is written as:

$$\left|\frac{\rho_2 + |\rho_3|e^{j\theta_3}e^{-2j\delta_2}}{1+\rho_2|\rho_3|e^{j\theta_3}e^{-2j\delta_2}}\right|^2 = \frac{\rho_2^2 + |\rho_3|^2 + 2\rho_2|\rho_3|\cos(2\delta_2-\theta_3)}{1+\rho_2^2|\rho_3|^2 + 2\rho_2|\rho_3|\cos(2\delta_2-\theta_3)} = \rho_1^2$$

Using the identity $\cos(2\delta_2 - \theta_3) = 2\cos^2(\delta_2 - \theta_3/2) - 1$, we find:

$$\begin{aligned}
\cos^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{\rho_1^2(1-\rho_2|\rho_3|)^2 - (\rho_2 - |\rho_3|)^2}{4\rho_2|\rho_3|(1-\rho_1^2)} \\
\sin^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{(\rho_2 + |\rho_3|)^2 - \rho_1^2(1+\rho_2|\rho_3|)^2}{4\rho_2|\rho_3|(1-\rho_1^2)}
\end{aligned} \tag{12.7.2}$$

Not every combination of ρ_1, ρ_2, ρ_3 will result into a solution for δ_2 because the left-hand sides must be positive and less than unity. If a solution for δ_2 exists, then δ_1 is determined from Eq. (12.7.1). Actually, there are two solutions for δ_2 corresponding to the ± signs of the square root of Eq. (12.7.2), that is, we have:

$$\delta_2 = \frac{1}{2}\theta_3 + \text{acos}\left[\pm\left(\frac{\rho_1^2(1-\rho_2|\rho_3|)^2 - (\rho_2 - |\rho_3|)^2}{4\rho_2|\rho_3|(1-\rho_1^2)}\right)^{1/2}\right] \tag{12.7.3}$$

If the resulting value of δ_2 is negative, it may be shifted by π or 2π to make it positive, and then solve for the electrical length $L_2 = \delta_2/2\pi$. An alternative way of writing Eqs. (12.7.2) is in terms of the segment impedances (see also Problem 6.6):

$$\begin{aligned}
\cos^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{(Z_2^2 - Z_3Z_0)(Z_3Z_1^2 - Z_0Z_2^2)}{Z_0(Z_2^2 - Z_3^2)(Z_1^2 - Z_2^2)} \\
\sin^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{Z_2^2(Z_0 - Z_3)(Z_1^2 - Z_0Z_3)}{Z_0(Z_2^2 - Z_3^2)(Z_1^2 - Z_2^2)}
\end{aligned} \tag{12.7.4}$$

where Z_3 is an equivalent "resistive" termination defined in terms of the load impedance through the relationship:

$$\frac{Z_3 - Z_2}{Z_3 + Z_2} = |\rho_3| = \left|\frac{Z_L - Z_2}{Z_L + Z_2}\right| \tag{12.7.5}$$

Clearly, if Z_L is real and greater than Z_2, then $Z_3 = Z_L$, whereas if it is less that Z_2, then, $Z_3 = Z_2^2/Z_L$. Eq. (12.7.4) shows more clearly the conditions for existence of solutions. In the special case when section-2 is a section of the main line, so that $Z_2 = Z_0$, then (12.7.4) simplifies to:

$$\begin{aligned}
\cos^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{Z_3Z_1^2 - Z_0^3}{(Z_3 + Z_0)(Z_1^2 - Z_0^2)} \\
\sin^2(\delta_2 - \frac{\theta_3}{2}) &= \frac{Z_0(Z_1^2 - Z_0Z_3)}{(Z_3 + Z_0)(Z_1^2 - Z_0^2)}
\end{aligned} \tag{12.7.6}$$

It is easily verified from these expressions that the condition for the existence of solutions is that the equivalent load impedance Z_3 lie within the intervals:

$$\begin{aligned}
\frac{Z_0^3}{Z_1^2} \le Z_3 \le \frac{Z_1^2}{Z_0}\,, &\quad \text{if} \quad Z_1 > Z_0 \\
\frac{Z_1^2}{Z_0} \le Z_3 \le \frac{Z_0^3}{Z_1^2}\,, &\quad \text{if} \quad Z_1 < Z_0
\end{aligned} \tag{12.7.7}$$

They may be combined into the single condition:

$$\boxed{\frac{Z_0}{S^2} \le Z_3 \le Z_0S^2}\,, \quad S = \frac{\max(Z_1, Z_0)}{\min(Z_1, Z_0)} = \text{swr}(Z_1, Z_0) \tag{12.7.8}$$

Example 12.7.1: *Matching range with 50- and 75-ohm lines.* If $Z_0 = 50$ and $Z_1 = 75$ ohm, then the following loads can be matched by this method:

$$\frac{50^3}{75^2} \le Z_3 \le \frac{75^2}{50} \quad \Rightarrow \quad 22.22 \le Z_3 \le 112.50\ \Omega$$

And, if $Z_0 = 75$ and $Z_1 = 50$, the following loads can be matched:

$$\frac{50^2}{75} \le Z_3 \le \frac{75^3}{50^2} \quad \Rightarrow \quad 33.33 \le Z_3 \le 168.75\ \Omega$$

In general, the farther Z_1 is from Z_0, the wider the range of loads that can be matched. For example, with $Z_0 = 75$ and $Z_1 = 300$ ohm, all loads in the range from 4.5 to 1200 ohm can be matched. □

The MATLAB function `twosect` implements the above design procedure. Its inputs are the impedances Z_0, Z_1, Z_2, and the complex Z_L, and its outputs are the two solutions for L_1 and L_2, if they exist. Its usage is as follows, where L_{12} is a 2×2 matrix whose rows are the two possible sets of values of L_1, L_2:

```
L12 = twosect(Z0,Z1,Z2,ZL);                % two-section series impedance transformer
```

The essential code in this function is as follows:

```
r1 = (Z1-Z0)/(Z1+Z0);
r2 = (Z2-Z1)/(Z2+Z1);
r3 = abs((ZL-Z2)/(ZL+Z2));
th3 = angle((ZL-Z2)/(ZL+Z2));

s = ((r2+r3)^2 - r1^2*(1+r2*r3)^2) / (4*r2*r3*(1-r1^2));
if (s<0)|(s>1), fprintf('no solution exists'); return; end

de2 = th3/2 + asin(sqrt(s)) * [1;-1];          % construct two solutions

G2 = (r2 + r3*exp(j*th3-2*j*de2)) ./ (1 + r2*r3*exp(j*th3-2*j*de2));

de1 = angle(-G2/r1)/2;

L1 = de1/2/pi; L2 = de2/2/pi;

L12 = mod([L1,L2], 0.5);                        % reduce modulo λ/2
```

Example 12.7.2: *Matching an antenna with coaxial cables.* A 29-MHz amateur radio antenna with input impedance of 38 ohm is to be fed by a 50-ohm RG-58/U cable. Design a two-section series impedance transformer consisting of a length of RG-59/U 75-ohm cable inserted into the main line at an appropriate distance from the antenna [943]. The velocity factor of both cables is 0.79.

Solution: Here, we have $Z_0 = 50$, $Z_1 = 75$, $Z_2 = Z_0$, and $Z_L = 38$ ohm. The call to the function `twosect` results in the MATLAB output for the electrical lengths of the segments:

$$L_{12} = \begin{bmatrix} 0.0536 & 0.3462 \\ 0.4464 & 0.1538 \end{bmatrix} \quad \Rightarrow \quad \begin{matrix} L_1 = 0.0536, & L_2 = 0.3462 \\ L_1 = 0.4464, & L_2 = 0.1538 \end{matrix}$$

Using the given velocity factor, the operating wavelength is $\lambda = 0.79\lambda_0 = 0.79c_0/f_0 = 8.1724$ m, where $f_0 = 29$ MHz. Therefore, the actual physical lengths for the segments are, for the first possible solution:

$$l_1 = 0.0536\lambda = 0.4379\text{ m} = 1.4367\text{ ft}\,, \qquad l_2 = 0.3462\lambda = 2.8290\text{ m} = 9.2813\text{ ft}$$

and for the second solution:

$$l_1 = 0.4464\lambda = 3.6483\text{ m} = 11.9695\text{ ft}\,, \quad l_2 = 0.1538\lambda = 1.2573\text{ m} = 4.1248\text{ ft}$$

Fig. 12.7.2 depicts the corresponding reflection responses at interface-1, $|\Gamma_1(f)|$, as a function of frequency. The standing wave ratio on the main line is also shown, that is, the quantity $S_1(f) = \big(1 + |\Gamma_1(f)|\big)/\big(1 - |\Gamma_1(f)|\big)$.

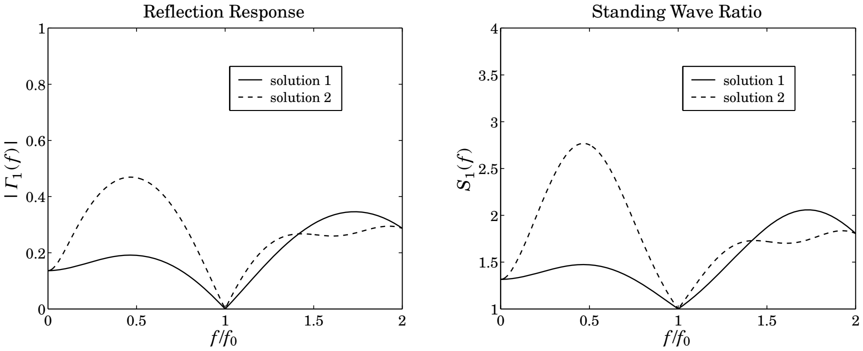


Fig. 12.7.2 Reflection response of two-section series transformer.

The reflection response was computed with the help of `multiline`. The typical MATLAB code for this example was:

```
Z0 = 50; Z1 = 75; ZL = 38;
c0 = 3e8; f0 = 29e6; vf = 0.79;
la0 = c0/f0; la = la0*vf;

L12 = twosect(Z0,Z1,Z0,ZL);

f = linspace(0,2,401);                               % in units of f0

G1 = abs(multiline([Z0,Z1,Z0],L12(1,:),ZL,f));      % reflection response 1
G2 = abs(multiline([Z0,Z1,Z0],L12(2,:),ZL,f));      % reflection response 2

S1=(1+G1)./(1-G1); S2=(1+G2)./(1-G2);               % SWRs
```

We note that the two solutions have unequal bandwidths. □

Example 12.7.3: *Matching a complex load.* Design a 75-ohm series section to be inserted into a 300-ohm line that feeds the load $600 + 900j$ ohm [943].

Solution: The MATLAB call

```
L12 = twosect(300, 75, 300, 600+900j);
```

produces the solutions: $L_1 = [0.3983,\ 0.1017]$ and $L_2 = [0.2420,\ 0.3318]$. □

One-section series impedance transformer

We mention briefly also the case of the one-section series impedance transformer, shown in Fig. 12.7.3. This is one of the earliest impedance transformers [937–941]. It has limited use in that not all complex loads can be matched, although its applicability can be extended somewhat [941].

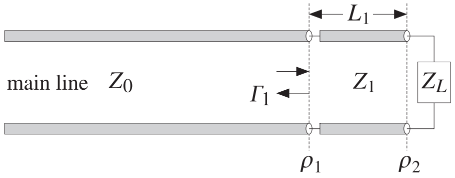


Fig. 12.7.3 One-section series impedance transformer.

Both the section impedance Z_1 and length L_1 are treated as unknowns to be fixed by requiring the matching condition $\Gamma_1 = 0$ at the operating frequency. It is left as an exercise (see Problem 12.9) to show that the solution is given by:

$$Z_1 = \sqrt{Z_0 R_L - \frac{Z_0 X_L^2}{Z_0 - R_L}}\,, \quad L_1 = \frac{1}{2\pi}\,\mathrm{atan}\left[\frac{Z_1(Z_0 - R_L)}{Z_0 X_L}\right] \tag{12.7.9}$$

provided that either of the following conditions is satisfied:

$$Z_0 < R_L \quad \text{or} \quad Z_0 > R_L + \frac{X_L^2}{R_L} \tag{12.7.10}$$

In particular, there is always a solution if Z_L is real. The MATLAB function `onesect` implements this method. It has usage:

```
[Z1,L1] = onesect(ZL,Z0);          % one-section series impedance transformer
```

where L_1 is the normalized length $L_1 = l_1/\lambda_1$, with l_1 and λ_1 the physical length and wavelength of the Z_1 section. The routine outputs the smallest positive L_1.

12.8 Single Stub Matching

Stub tuners are widely used to match any complex load[†] to a main line. They consist of shorted or opened segments of the line, connected in parallel or in series with the line at a appropriate distances from the load.

[†] The resistive part of the load must be non-zero. Purely reactive loads cannot be matched to a real line impedance by this method nor by any of the other methods discussed in this chapter. This so because the transformation of a reactive load through the matching circuits remains reactive.

In coaxial cable or two-wire line applications, the stubs are obtained by cutting appropriate lengths of the main line. Shorted stubs are usually preferred because opened stubs may radiate from their opened ends. However, in microwave integrated circuits employing microstrip lines, radiation is not as a major concern because of their smaller size, and either opened or shorted stubs may be used.

The single stub tuner is perhaps the most widely used matching circuit and can match any load. However, it is sometimes inconvenient to connect to the main line if different loads are to be matched. In such cases, double stubs may be used, but they cannot match all loads. Triple stubs can match any load. A single stub tuner is shown in Figs. 12.8.1 and 12.8.2, connected in parallel and in series.

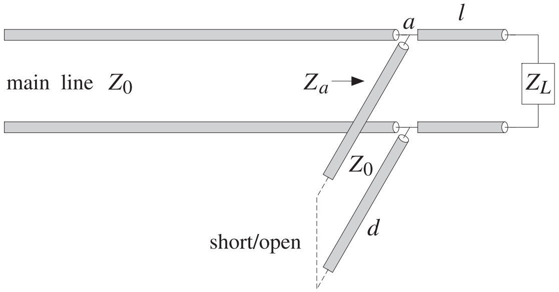


Fig. 12.8.1 Parallel connection of single stub tuner.

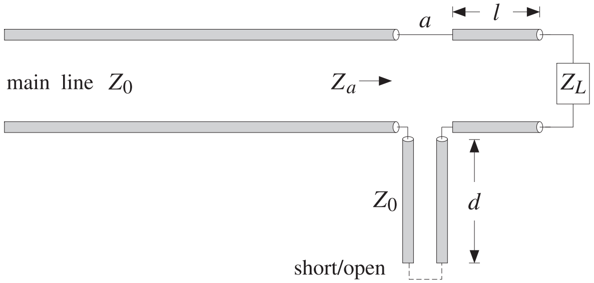


Fig. 12.8.2 Series connection of single stub tuner.

In the parallel case, the admittance $Y_a = 1/Z_a$ at the stub location a is the sum of the admittances of the length-d stub and the wave admittance at distance l from the load, that is,

$$Y_a = Y_l + Y_{\text{stub}} = Y_0\frac{1-\Gamma_l}{1+\Gamma_l} + Y_{\text{stub}}$$

where $\Gamma_l = \Gamma_L e^{-2j\beta l}$. The admittance of a short-circuited stub is $Y_{\text{stub}} = -jY_0\cot\beta d$, and of an open-circuited one, $Y_{\text{stub}} = jY_0\tan\beta d$. The matching condition is that $Y_a = Y_0$. Assuming a short-circuited stub, we have:

$$Y_0\frac{1-\Gamma_l}{1+\Gamma_l} - jY_0\cot\beta d = Y_0 \quad\Rightarrow\quad \frac{1-\Gamma_l}{1+\Gamma_l} - j\cot\beta d = 1$$

which can be rearranged into the form:

$$2j\tan\beta d = 1 + \frac{1}{\Gamma_l} \qquad (12.8.1)$$

Inserting $\Gamma_l = \Gamma_L e^{-2j\beta l} = |\Gamma_L|e^{j\theta_L - 2j\beta l}$, where $\Gamma_L = |\Gamma_L|e^{j\theta_L}$ is the polar form of the load reflection coefficient, we may write (12.8.1) as:

$$2j\tan\beta d = 1 + \frac{e^{j(2\beta l - \theta_L)}}{|\Gamma_L|} \qquad (12.8.2)$$

Equating real and imaginary parts, we obtain the equivalent conditions:

$$\cos(2\beta l - \theta_L) = -|\Gamma_L|\,, \quad \tan\beta d = \frac{\sin(2\beta l - \theta_L)}{2|\Gamma_L|} = -\frac{1}{2}\tan(2\beta l - \theta_L) \qquad (12.8.3)$$

The first of (12.8.3) may be solved resulting in two solutions for l; then, the second equation may be solved for the corresponding values of d:

$$\beta l = \frac{1}{2}\theta_L \pm \frac{1}{2}\,\mathrm{acos}\big(-|\Gamma_L|\big)\,, \quad \beta d = \mathrm{atan}\big(-\frac{1}{2}\tan(2\beta l - \theta_L)\big) \qquad (12.8.4)$$

The resulting values of l, d must be made positive by reducing them modulo $\lambda/2$. In the case of an open-circuited shunt stub, the first equation in (12.8.3) remains the same, and in the second we must replace $\tan\beta d$ by $-\cot\beta d$. In the series connection of a shorted stub, the impedances are additive at point a, resulting in the condition:

$$Z_a = Z_l + Z_{\text{stub}} = Z_0\frac{1+\Gamma_l}{1-\Gamma_l} + jZ_0\tan\beta d = Z_0 \quad\Rightarrow\quad \frac{1+\Gamma_l}{1-\Gamma_l} - \tan\beta d = 1$$

This may be solved in a similar fashion as Eq. (12.8.1). We summarize below the solutions in the four cases of parallel or series connections with shorted or opened stubs:

$$\begin{array}{lll}
\beta l = \frac{1}{2}\big[\theta_L \pm \mathrm{acos}(-|\Gamma_L|)\big], & \beta d = \mathrm{atan}\big(-\frac{1}{2}\tan(2\beta l - \theta_L)\big), & \text{parallel/shorted} \\
\beta l = \frac{1}{2}\big[\theta_L \pm \mathrm{acos}(-|\Gamma_L|)\big], & \beta d = \mathrm{acot}\big(\frac{1}{2}\tan(2\beta l - \theta_L)\big), & \text{parallel/opened} \\
\beta l = \frac{1}{2}\big[\theta_L \pm \mathrm{acos}(|\Gamma_L|)\big], & \beta d = \mathrm{acot}\big(\frac{1}{2}\tan(2\beta l - \theta_L)\big), & \text{series/shorted} \\
\beta l = \frac{1}{2}\big[\theta_L \pm \mathrm{acos}(|\Gamma_L|)\big], & \beta d = \mathrm{atan}\big(-\frac{1}{2}\tan(2\beta l - \theta_L)\big), & \text{series/opened}
\end{array}$$

The MATLAB function `stub1` implements these equations. Its input is the normalized load impedance, $z_L = Z_L/Z_0$, and the desired type of stub. Its outputs are the dual solutions for the lengths d, l, arranged in the rows of a 2x2 matrix `dl`. Its usage is as follows:

```
dl = stub1(zL,type);                  % single stub tuner
```

The parameter `type` takes on the string values 'ps', 'po', 'ss', 'so', for parallel/short, parallel/open, series/short, series/open stubs.

Example 12.8.1: The load impedance $Z_L = 10 - 5j$ ohm is to be matched to a 50-ohm line. The normalized load is $z_L = Z_L/Z_0 = 0.2 - 0.1j$. The MATLAB calls, `dl=stub1(zL,type)`, result into the following solutions for the cases of parallel/short, parallel/open, series/short, series/open stubs:

$$\begin{bmatrix} 0.0806 & 0.4499 \\ 0.4194 & 0.0831 \end{bmatrix}, \begin{bmatrix} 0.3306 & 0.4499 \\ 0.1694 & 0.0831 \end{bmatrix}, \begin{bmatrix} 0.1694 & 0.3331 \\ 0.3306 & 0.1999 \end{bmatrix}, \begin{bmatrix} 0.4194 & 0.3331 \\ 0.0806 & 0.1999 \end{bmatrix}$$

Each row represents a possible solution for the electrical lengths d/λ and l/λ. We illustrate below the solution details for the parallel/short case.

Given the load impedance $z_L = 0.2 - 0.1j$, we calculate the reflection coefficient and put it in polar form:

$$\Gamma_L = \frac{z_L - 1}{z_L + 1} = -0.6552 - 0.1379j \quad\Rightarrow\quad |\Gamma_L| = 0.6695\,, \quad \theta_L = -2.9341 \text{ rad}$$

Then, the solution of Eq. (12.8.4) is:

$$\beta l = \frac{1}{2}\big[\theta_L \pm \mathrm{acos}(-|\Gamma_L|)\big] = \frac{1}{2}\big[-2.9341 \pm \mathrm{acos}(-0.6695)\big] = \frac{1}{2}\big[-2.9341 \pm 2.3044\big]$$

which gives the two solutions:

$$\beta l = \frac{2\pi l}{\lambda} = \begin{bmatrix} -0.3149 \text{ rad} \\ -2.6192 \text{ rad} \end{bmatrix} \quad\Rightarrow\quad l = \frac{\lambda}{2\pi}\begin{bmatrix} -0.3149 \\ -2.6192 \end{bmatrix} = \begin{bmatrix} -0.0501\lambda \\ -0.4169\lambda \end{bmatrix}$$

These may be brought into the interval $[0, \lambda/2]$ by adding enough multiples of $\lambda/2$. The built-in MATLAB function `mod` does just that. In this case, a single multiple of $\lambda/2$ suffices, resulting in:

$$l = \begin{bmatrix} -0.0501\lambda + 0.5\lambda \\ -0.4169\lambda + 0.5\lambda \end{bmatrix} = \begin{bmatrix} 0.4499\lambda \\ 0.0831\lambda \end{bmatrix} \quad\Rightarrow\quad \beta l = \begin{bmatrix} 2.8267 \text{ rad} \\ 0.5224 \text{ rad} \end{bmatrix}$$

With these values of βl, we calculate the stub length d:

$$\beta d = \mathrm{atan}\big(-\frac{1}{2}\tan(2\beta l - \theta_L)\big) = \begin{bmatrix} 0.5064 \text{ rad} \\ -0.5064 \text{ rad} \end{bmatrix} \quad\Rightarrow\quad d = \begin{bmatrix} 0.0806\lambda \\ -0.0806\lambda \end{bmatrix}$$

Shifting the second d by $\lambda/2$, we finally find:

$$d = \begin{bmatrix} 0.0806\lambda \\ -0.0806\lambda + 0.5\lambda \end{bmatrix} = \begin{bmatrix} 0.0806\lambda \\ 0.4194\lambda \end{bmatrix}, \quad \beta d = \begin{bmatrix} 0.5064 \text{ rad} \\ 2.6351 \text{ rad} \end{bmatrix}$$

Next, we verify the matching condition. The load admittance is $y_L = 1/z_L = 4 + 2j$. Propagating it to the left of the load by a distance l, we find for the two values of l and for the corresponding values of d:

$$y_l = \frac{y_L + j\tan\beta l}{1 + jy_L\tan\beta l} = \begin{bmatrix} 1.0000 + 1.8028j \\ 1.0000 - 1.8028j \end{bmatrix}, \quad y_{\text{stub}} = -j\cot\beta d = \begin{bmatrix} -1.8028j \\ 1.8028j \end{bmatrix}$$

For both solutions, the susceptance of y_l is canceled by the susceptance of the stub, resulting in the matched total normalized admittance $y_a = y_l + y_{\text{stub}} = 1$. □

Example 12.8.2: Match the antenna and feed line of Example 12.7.2 using a single shorted or opened stub. Plot the corresponding matched reflection responses.

Solution: The normalized load impedance is $z_L = 38/50 = 0.76$. The MATLAB function to `stub1` yields the following solutions for the lengths d, l, in the cases of parallel/short, parallel/open, series/short, series/open stubs:

$$\begin{bmatrix} 0.2072 & 0.3859 \\ 0.2928 & 0.1141 \end{bmatrix}, \begin{bmatrix} 0.4572 & 0.3859 \\ 0.0428 & 0.1141 \end{bmatrix}, \begin{bmatrix} 0.0428 & 0.3641 \\ 0.4572 & 0.1359 \end{bmatrix}, \begin{bmatrix} 0.2928 & 0.3641 \\ 0.2072 & 0.1359 \end{bmatrix},$$

These numbers must be multiplied by λ_0, the free-space wavelength corresponding to the operating frequency of $f_0 = 29$ MHz. The resulting reflection responses $|\Gamma_a(f)|$ at the connection point a of the stub, corresponding to all the pairs of d, l are shown in Fig. 12.8.3. For example, in the parallel/short case, Γ_a is calculated by

$$\Gamma_a = \frac{1 - y_a}{1 + y_a}, \quad y_a = \frac{1 - \Gamma_L e^{-2j\beta l}}{1 + \Gamma_L e^{-2j\beta l}} - j\cot\beta d, \quad \beta l = 2\pi\frac{f}{f_0}\frac{l}{\lambda_0}, \quad \beta d = 2\pi\frac{f}{f_0}\frac{d}{\lambda_0}$$

We note that different solutions can have very different bandwidths. □

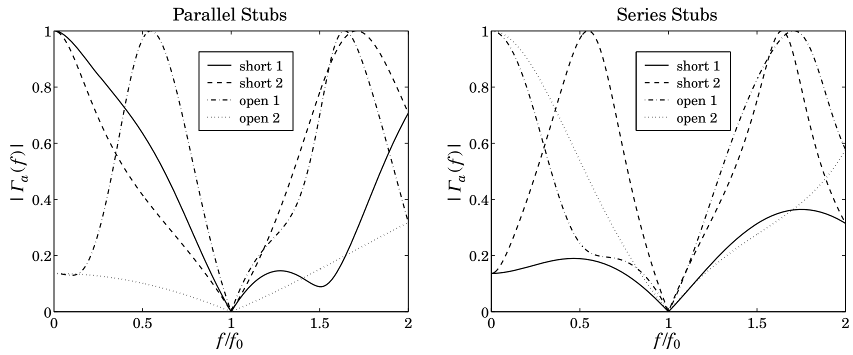


Fig. 12.8.3 Reflection response of single stub matching solutions.

12.9 Balanced Stubs

In microstrip realizations of single-stub tuners, balanced stubs are often used to reduce the transitions between the series and shunt segments. Fig. 12.9.1 depicts two identical balanced stubs connected at opposite sides of the main line.

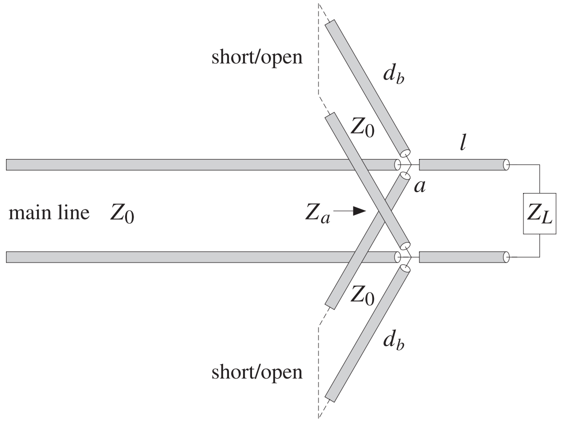


Fig. 12.9.1 Balanced stubs.

Because of the parallel connection, the total admittance of the stubs will be double that of each leg, that is, $Y_{\text{bal}} = 2Y_{\text{stub}}$. A single unbalanced stub of length d can be converted into an equivalent balanced stub of length d_b by requiring that the two configurations provide the same admittance. Depending on whether shorted or opened stubs are used, we obtain the relationships between d_b and d:

$$\begin{aligned} 2\cot\beta d_b = \cot\beta d \quad &\Rightarrow \quad d_b = \frac{\lambda}{2\pi}\,\text{acot}\big(0.5\cot\beta d\big) \qquad \text{(shorted)} \\ 2\tan\beta d_b = \tan\beta d \quad &\Rightarrow \quad d_b = \frac{\lambda}{2\pi}\,\text{atan}\big(0.5\tan\beta d\big) \qquad \text{(opened)} \end{aligned} \tag{12.9.1}$$

The microstrip realization of such a balanced stub is shown in Fig. 12.9.2. The figure also shows the use of balanced stubs for quarter-wavelength transformers with a shunt stub as discussed in Sec. 12.6.

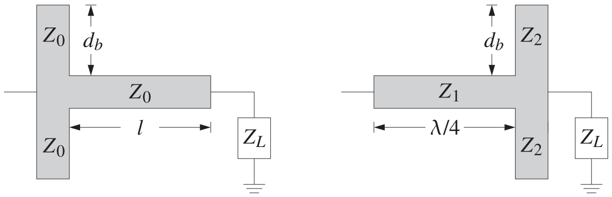


Fig. 12.9.2 Balanced microstrip single-stub and quarter-wavelength transformers.

If the shunt stub has length $\lambda/8$ or $3\lambda/8$, then the impedance Z_2 of each leg must be double that of the single-stub case. On the other hand, if the impedance Z_2 is fixed, then the stub length d_b of each leg may be calculated by Eq. (12.9.1).

12.10 Double and Triple Stub Matching

Because the stub distance l from the load depends on the load impedance to be matched, the single-stub tuner is inconvenient if several different load impedances are to be matched, each requiring a different value for l.

The double-stub tuner, shown in Fig. 12.10.1, provides an alternative matching method in which two stubs are used, one at the load and another at a *fixed* distance l from the load, where typically, $l = \lambda/8$. Only the stub lengths d_1, d_2 need to be adjusted to match the load impedance.

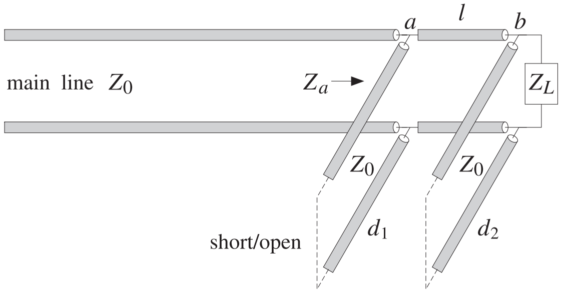


Fig. 12.10.1 Double stub tuner.

The two stubs are connected in parallel to the main line and can be short- or open-circuited. We discuss the matching conditions for the case of shorted stubs.

Let $Y_L = 1/Z_L = G_L + jB_L$ be the load admittance, and define its normalized version $y_L = Y_L/Y_0 = g_L + jb_L$, where g_L, b_L are the normalized load conductance and susceptance. At the connection points a, b, the total admittance is the sum of the wave admittance of the line and the stub admittance:

$$y_a = y_l + y_{\text{stub},1} = \frac{y_b + j\tan\beta l}{1 + jy_b\tan\beta l} - j\cot\beta d_1$$

$$y_b = y_L + y_{\text{stub},2} = g_L + j(b_L - \cot\beta d_2)$$

The matching condition is $y_a = 1$, which gives rise to two equations that can be solved for the unknown lengths d_1, d_2. It is left as an exercise (see Problem 12.10) to show that the solutions are given by:

$$\cot\beta d_2 = b_L - b\,, \quad \cot\beta d_1 = \frac{1 - b\tan\beta l - g_L}{g_L\tan\beta l} \tag{12.10.1}$$

where

$$b = \cot\beta l \pm \sqrt{g_L(g_{\max} - g_L)}\,, \quad g_{\max} = 1 + \cot^2\beta l = \frac{1}{\sin^2\beta l} \tag{12.10.2}$$

Evidently, the condition for the existence of a real-valued b is that the load conductance g_L be less than $g_{\max}$, that is, $g_L \le g_{\max}$. If this condition is not satisfied, the load cannot be matched with any stub lengths d_1, d_2. Stub separations near $\lambda/2$, or near zero, result in $g_{\max} = \infty$, but are not recommended because they have very narrow bandwidths [846].

Assuming $l \le \lambda/4$, the condition $g_L \le g_{\max}$ can be turned around into a condition for the maximum length l that will admit a matching solution for the given load:

$$l \le l_{\max} = \frac{\lambda}{2\pi}\,\text{asin}\Big(\frac{1}{\sqrt{g_L}}\Big) \qquad \text{(maximum stub separation)} \tag{12.10.3}$$

If the existence condition is satisfied, then Eq. (12.10.2) results in two solutions for b and, hence for, d_1, d_2. The lengths d_1, d_2 must be reduced modulo $\lambda/2$ to bring them within the minimum interval $[0, \lambda/2]$.

If any of the stubs are open-circuited, the corresponding quantity $\cot\beta d_i$ must be replaced by $-\tan\beta d_i = \cot(\beta d_i - \pi/2)$.

The MATLAB function `stub2` implements the above design procedure. Its inputs are the normalized load impedance $z_L = Z_L/Z_0$, the stub separation l, and the stub types, and its outputs are the two possible solutions for the d_1, d_2. Its usage is as follows:

```
d12 = stub2(zL,l,type);          % double stub tuner
d12 = stub2(zL,l);               % equivalent to type='ss'
d12 = stub2(zL);                 % equivalent to l = 1/8 and type='ss'
```

The parameter `type` takes on the strings values: 'ss', 'so', 'os', 'oo', for short/short, short/open, open/short, open/open stubs. If the existence condition fails, the function outputs the maximum separation $l_{\max}$ that will admit a solution.

A triple stub tuner, shown in Fig. 12.10.2, can match any load. The distances l_1, l_2 between the stubs are fixed and only the stub lengths d_1, d_2, d_3 are adjustable.

The first two stubs (from the left) can be thought of as a double-stub tuner. The purpose of the third stub at the load is to ensure that the wave impedance seen by the double-stub tuner satisfies the existence condition $g_L \le g_{\max}$.

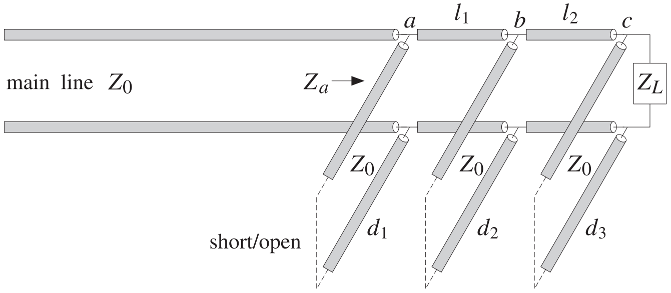


Fig. 12.10.2 Triple stub tuner.

The total admittance at the load point c, and its propagated version by distance l_2 to point b are given by:

$$y_l = \frac{y_c + j\tan\beta l_2}{1 + jy_c\tan\beta l_2}\,, \quad y_c = y_L + y_{\text{stub},3} = g_L + jb_L - j\cot\beta d_3 = g_L + jb \tag{12.10.4}$$

where $b = b_L - \cot\beta d_3$. The corresponding conductance is:

$$g_l = \mathrm{Re}(y_l) = \frac{g_L(1+\tan^2\beta l_2)}{(b\tan\beta l_2 - 1)^2 + g_L^2\tan^2\beta l_2} \tag{12.10.5}$$

The first two stubs see the effective load y_l. The double-stub problem will have a solution provided $g_l \le g_{\max,1} = 1/\sin^2\beta l_1$. The length d_3 of the third stub is adjusted to ensure this condition. To parametrize the possible solutions, we introduce a "smallness" parameter $e < 1$ such that $g_l = e g_{\max,1}$. This gives the existence condition:

$$g_l = \frac{g_L(1+\tan^2\beta l_2)}{(b\tan\beta l_2 - 1)^2 + g_L^2\tan^2\beta l_2} = e g_{\max,1}$$

which can be rewritten in the form:

$$(b - \cot\beta l_2)^2 = g_L(g_{\max,2} - e g_{\max,1} g_L) = g_L^2 g_{\max,1}(e_{\max} - e)$$

where we defined $g_{\max,2} = 1 + \cot^2\beta l_2 = 1/\sin^2\beta l_2$ and $e_{\max} = g_{\max,2}/(g_L g_{\max,1})$. If $e_{\max} < 1$, we may replace e by the minimum of the chosen e and $e_{\max}$. But if $e_{\max} > 1$, we just use the chosen e. In other words, we replace the above condition with:

$$(b - \cot\beta l_2)^2 = g_L^2 g_{\max,1}(e_{\max} - e_{\min})\,, \quad e_{\min} = \min(e, e_{\max}) \tag{12.10.6}$$

It corresponds to setting $g_l = e_{\min} g_{\max,1}$. Solving Eq. (12.10.6) for $\cot\beta d_3$ gives the two solutions:

$$\cot\beta d_3 = b_L - b\,, \quad b = \cot\beta l_2 \pm g_L\sqrt{g_{\max,1}(e_{\max} - e_{\min})} \tag{12.10.7}$$

For each of the two values of d_3, there will be a feasible solution to the double-stub problem, which will generate two possible solutions for d_1, d_2. Thus, there will be a total of four triples d_1, d_2, d_3 that will satisfy the matching conditions. Each stub can be shorted or opened, resulting into eight possible choices for the stub triples.

The MATLAB function `stub3` implements the above design procedure. It generates a 4×3 matrix of solutions and its usage is:

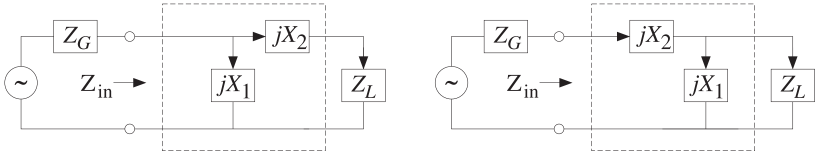
```
d123 = stub3(zL,l1,l2,type,e);      % triple stub tuner
d123 = stub3(zL,l1,l2,type);        % equivalent to e = 0.9
d123 = stub3(zL,l1,l2);             % equivalent to e = 0.9, type='sss'
d123 = stub3(zL);                   % equivalent to e = 0.9, type='sss', l1 = l2 = 1/8
```

where `type` takes on one of the eight possible string values, defining whether the first, second, or third stubs are short- or open-circuited: 'sss', 'sso', 'sos', 'soo', 'oss', 'oso', 'oos', 'ooo'.

12.11 *L*-Section Lumped Reactive Matching Networks

Impedance matching by stubs or series transmission line segments is appropriate at higher frequencies, such as microwave frequencies. At lower RF frequencies, lumped-parameter circuit elements may be used to construct a matching network. Here, we discuss *L*-section, *Π*-section, and *T*-section matching networks.

The *L*-section matching network shown in Fig. 12.11.1 uses only reactive elements (inductors or capacitors) to conjugately match any load impedance Z_L to any generator impedance Z_G. The use of reactive elements minimizes power losses in the matching network.



normal *L*-section ($R_G > R_L$) reversed *L*-section ($R_G < R_L$)

Fig. 12.11.1 *L*-section reactive conjugate matching network.

L-section networks are used to match the input and output impedances of amplifier circuits [982–990] and also to match transmitters to feed lines [43,44,944–951].

An arbitrary load impedance may be matched by a normal *L*-section, or if that is not possible, by a reversed *L*-section. Sometimes both normal and reversed types are possible. We derive below the conditions for the existence of a matching solution of a particular type.

The inputs to the design procedure are the complex load and generator impedances $Z_L = R_L + jX_L$ and $Z_G = R_G + jX_G$. The outputs are the reactances X_1, X_2. For either type, the matching network transforms the load impedance Z_L into the complex conjugate of the generator impedance, that is,

$$\boxed{Z_{\mathrm{in}} = Z_G^*} \qquad \text{(conjugate match)} \tag{12.11.1}$$

where Z_{in} is the input impedance looking into the *L*-section:

$$\begin{aligned} Z_{\mathrm{in}} &= \frac{Z_1(Z_2+Z_L)}{Z_1+Z_2+Z_L} \qquad \text{(normal)} \\ Z_{\mathrm{in}} &= Z_2 + \frac{Z_1 Z_L}{Z_1+Z_L} \qquad \text{(reversed)} \end{aligned} \tag{12.11.2}$$

with $Z_1 = jX_1$ and $Z_2 = jX_2$. Inserting Eqs. (12.11.2) into the condition (12.11.1) and equating the real and imaginary parts of the two sides, we obtain a system of equations for X_1, X_2 with solutions for the two types:

$$\boxed{\begin{aligned} X_1 &= \frac{X_G \pm R_G Q}{\dfrac{R_G}{R_L} - 1} \\ X_2 &= -(X_L \pm R_L Q) \\ Q &= \sqrt{\frac{R_G}{R_L} - 1 + \frac{X_G^2}{R_G R_L}} \end{aligned}}\ \text{(normal)}\,, \quad \boxed{\begin{aligned} X_1 &= \frac{X_L \pm R_L Q}{\dfrac{R_L}{R_G} - 1} \\ X_2 &= -(X_G \pm R_G Q) \\ Q &= \sqrt{\frac{R_L}{R_G} - 1 + \frac{X_L^2}{R_G R_L}} \end{aligned}}\ \text{(reversed)} \tag{12.11.3}$$

If the load and generator impedances are both resistive, so that $X_L = 0$ and $X_G = 0$, the above solutions take the particularly simple forms:

$$\boxed{\begin{aligned} X_1 &= \pm\frac{R_G}{Q} \\ X_2 &= \mp R_L Q \\ Q &= \sqrt{\frac{R_G}{R_L} - 1} \end{aligned}} \quad \text{(normal)}, \qquad \boxed{\begin{aligned} X_1 &= \pm\frac{R_L}{Q} \\ X_2 &= \mp R_G Q \\ Q &= \sqrt{\frac{R_L}{R_G} - 1} \end{aligned}} \quad \text{(reversed)} \tag{12.11.4}$$

We note that the reversed solution is obtained from the normal one by exchanging Z_L with Z_G. Both solution types assume that $R_G \neq R_L$. If $R_G = R_L$, then for either type, we have the solution:

$$X_1 = \infty, \quad X_2 = -(X_L + X_G) \tag{12.11.5}$$

Thus, X_1 is open-circuited and X_2 is such that $X_2 + X_L = -X_G$. The Q quantities play the role of series impedance Q-factors. Indeed, the X_2 equations in all cases imply that Q is equal to the ratio of the total series reactance by the corresponding series resistance, that is, $(X_2 + X_L)/R_L$ or $(X_2 + X_G)/R_G$.

The conditions for real-valued solutions for X_1, X_2 are that the Q factors in (12.11.3) and (12.11.4) be real-valued or that the quantities under their square roots be non-negative. When $R_L \neq R_G$, it is straightforward to verify that this happens in the following four mutually exclusive cases:

existence conditions	L-section types
$R_G > R_L\,, \quad \lvert X_L\rvert \geq \sqrt{R_L(R_G - R_L)}$	normal and reversed
$R_G > R_L\,, \quad \lvert X_L\rvert < \sqrt{R_L(R_G - R_L)}$	normal only
$R_G < R_L\,, \quad \lvert X_G\rvert \geq \sqrt{R_G(R_L - R_G)}$	normal and reversed
$R_G < R_L\,, \quad \lvert X_G\rvert < \sqrt{R_G(R_L - R_G)}$	reversed only

(12.11.6)

It is evident that a solution of one or the other type always exists. When $R_G > R_L$ a normal section always exists, and when $R_G < R_L$ a reversed one exists. The MATLAB function `lmatch` implements Eqs. (12.11.3). Its usage is as follows:

```
X12 = lmatch(ZG,ZL,type);                    % L-section matching
```

where `type` takes on the string values `'n'` or `'r'` for a normal or reversed L-section. The two possible solutions for X_1, X_2 are returned in the rows of the 2×2 matrix X_{12}.

Example 12.11.1: Design an L-section matching network for the conjugate match of the load impedance $Z_L = 100+50j$ ohm to the generator $Z_G = 50+10j$ ohm at 500 MHz. Determine the capacitance or inductance values for the matching network.

Solution: The given impedances satisfy the last of the four conditions of Eq. (12.11.6). Therefore, only a reversed L-section will exist. Its two solutions are:

$$X_{12} = \text{lmatch}(50 + 10j, 100 + 50j, \texttt{'r'}) = \begin{bmatrix} 172.4745 & -71.2372 \\ -72.4745 & 51.2372 \end{bmatrix}$$

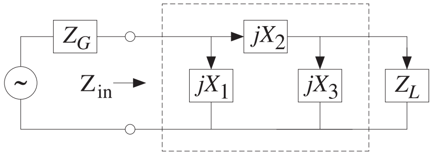
The first solution has a capacitive $X_2 = -71.2372$ and an inductive $X_1 = 172.4745$. Setting $X_2 = 1/j\omega C$ and $X_1 = j\omega L$, where $\omega = 2\pi f = 2\pi 500 \cdot 10^6$ rad/sec, we determine the corresponding values of C and L to be $C = 4.47$ pF and $L = 54.90$ nH.

The second solution has an inductive $X_2 = 51.2372$ and a capacitive $X_1 = -72.4745$. Setting $X_2 = j\omega L$ and $X_1 = 1/j\omega C$, we find in this case, $L = 16.3$ nH and $C = 4.39$ pF. Of the two solutions, the one with the smaller values is generally preferred. □

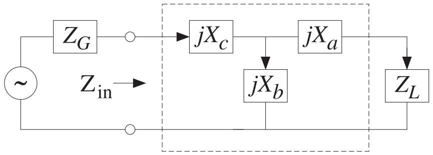
12.12 Pi-Section Lumped Reactive Matching Networks

Although the L-section network can match an arbitrary load to an arbitrary source, its bandwidth and Q-factor are fixed uniquely by the values of the load and source impedances through Eqs. (12.11.3).

The Π-section network, shown together with its T-section equivalent in Fig. 12.12.1, has an extra degree of freedom that allows one to control the bandwidth of the match. In particular, the bandwidth can be made as narrow as desired.



Π-section



T-section

Fig. 12.12.1 Π- and T-section matching networks.

The Π, T networks (also called Δ, Y networks) can be transformed into each other by the following standard impedance transformations, which are cyclic permutations of each other:

$$\begin{aligned} Z_a &= \frac{Z_2 Z_3}{U}, \quad Z_b = \frac{Z_3 Z_1}{U}, \quad Z_c = \frac{Z_1 Z_2}{U}, \qquad U = Z_1 + Z_2 + Z_3 \\ Z_1 &= \frac{V}{Z_a}, \quad Z_2 = \frac{V}{Z_b}, \quad Z_3 = \frac{V}{Z_c}, \qquad V = Z_a Z_b + Z_b Z_c + Z_c Z_a \end{aligned} \tag{12.12.1}$$

Because Z_1, Z_2, Z_3 are purely reactive, $Z_1 = jX_1$, $Z_2 = jX_2$, $Z_3 = jX_3$, so will be Z_a, Z_b, Z_c, with $Z_a = jX_a$, $Z_b = jX_b$, $Z_c = jX_c$.

The MATLAB functions `pi2t` and `t2pi` transform between the two parameter sets. The function `pi2t` takes in the array of three values $Z_{123} = [Z_1, Z_2, Z_3]$ and outputs $Z_{abc} = [Z_a, Z_b, Z_c]$, and `t2pi` does the reverse. Their usage is:

```
Zabc = pi2t(Z123);            % Π to T transformation
Z123 = t2pi(Zabc);            % T to Π transformation
```

One of the advantages of T networks is that often they result in more practical values for the circuit elements; however, they tend to be more lossy [43,44].

Here we discuss only the design of the Π matching network. It can be transformed into a T network if so desired. Fig. 12.12.2 shows the design procedure, in which the Π network can be thought of as two L-sections arranged back to back, by splitting the series reactance X_2 into two parts, $X_2 = X_4 + X_5$.

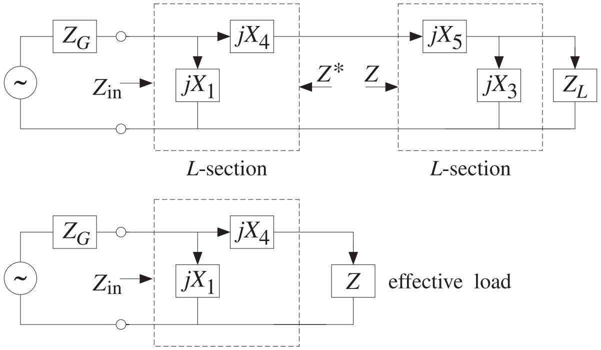


Fig. 12.12.2 Equivalent L-section networks.

An additional degree of freedom is introduced into the design by an intermediate reference impedance, say $Z = R + jX$, such that looking into the right L-section the input impedance is Z, and looking into the left L-section, it is Z^*.

Denoting the L-section impedances by $Z_1 = jX_1$, $Z_4 = jX_4$ and $Z_3 = jX_3$, $Z_5 = jX_5$, we have the conditions:

$$Z_{\rm left} = Z_4 + \frac{Z_1 Z_G}{Z_1 + Z_G} = Z^*\,, \qquad Z_{\rm right} = Z_5 + \frac{Z_3 Z_L}{Z_3 + Z_L} = Z \tag{12.12.2}$$

As shown in Fig. 12.12.2, the right L-section and the load can be replaced by the effective load impedance $Z_{\rm right} = Z$. Because Z_1 and Z_4 are purely reactive, their conjugates will be $Z_1^* = -Z_1$ and $Z_4^* = -Z_4$. It then follows that the first of Eqs. (12.12.2) can be rewritten as the equivalent condition:

$$Z_{\rm in} = \frac{Z_1(Z_4 + Z)}{Z_1 + Z_4 + Z} = Z_G^* \tag{12.12.3}$$

This is precisely the desired conjugate matching condition that must be satisfied by the network (as terminated by the effective load Z.)

Eq. (12.12.3) can be interpreted as the result of matching the source Z_G to the load Z with a normal L-section. An equivalent point of view is to interpreted the first of Eqs. (12.12.2) as the result of matching the source Z to the load Z_G using a reversed L-section.

Similarly, the second of Eqs. (12.12.2) is the result of matching the source Z^* to the load Z_L (because the input impedance looking into the right section is then $(Z^*)^* = Z$.) Thus, the reactances of the two L-sections can be obtained by the two successive calls to `lmatch`:

$$\begin{aligned} X_{14} &= [X_1, X_4] = \text{lmatch}(Z_G, Z, \texttt{'n'}) = \text{lmatch}(Z, Z_G, \texttt{'r'}) \\ X_{35} &= [X_3, X_5] = \text{lmatch}(Z^*, Z_L, \texttt{'r'}) \end{aligned} \tag{12.12.4}$$

In order for Eqs. (12.12.4) to always have a solution, the resistive part of Z must satisfy the conditions (12.11.6). Thus, we must choose $R < R_G$ and $R < R_L$, or equivalently:

$$R < R_{\rm min}\,, \qquad R_{\rm min} = \min(R_G, R_L) \tag{12.12.5}$$

Otherwise, Z is arbitrary. For design purposes, the nominal Q factors of the left and right sections can be taken to be the quantities:

$$Q_G = \sqrt{\frac{R_G}{R} - 1}\,, \quad Q_L = \sqrt{\frac{R_L}{R} - 1} \tag{12.12.6}$$

The maximum of the two is the one with the maximum value of R_G or R_L, that is,

$$\boxed{Q = \sqrt{\frac{R_{\rm max}}{R} - 1}}\,, \qquad R_{\rm max} = \max(R_G, R_L) \tag{12.12.7}$$

This Q-factor can be thought of as a parameter that controls the bandwidth. Given a value of Q, the corresponding R is obtained by:

$$\boxed{R = \frac{R_{\rm max}}{Q^2 + 1}} \tag{12.12.8}$$

For later reference, we may express Q_G, Q_L in terms of Q as follows:

$$Q_G = \sqrt{\frac{R_G}{R_{\rm max}}(Q^2+1) - 1}\,, \quad Q_L = \sqrt{\frac{R_L}{R_{\rm max}}(Q^2+1) - 1} \tag{12.12.9}$$

Clearly, one or the other of Q_L, Q_G is equal to Q. We note also that Q may not be less than the value $Q_{\rm min}$ achievable by a *single* L-section match. This follows from the equivalent conditions:

$$\boxed{Q > Q_{\rm min} \quad \Leftrightarrow \quad R < R_{\rm min}}\,, \qquad Q_{\rm min} = \sqrt{\frac{R_{\rm max}}{R_{\rm min}} - 1} \tag{12.12.10}$$

The MATLAB function `pmatch` implements the design equations (12.12.4) and then constructs $X_2 = X_4 + X_5$. Because there are two solutions for X_4 and two for X_5, we can add them in four different ways, leading to four possible solutions for the reactances of the Π network.

The inputs to `pmatch` are the impedances Z_G, Z_L and the reference impedance Z, which must satisfy the condition (12.12.10). The output is a 4×3 matrix X_{123} whose rows are the different solutions for X_1, X_2, X_3:

```
X123 = pmatch(ZG,ZL,Z);            % Π matching network design
```

The analytical form of the solutions can be obtained easily by applying Eqs. (12.11.3) to the two cases of Eq. (12.12.4). In particular, if the load and generator impedances are real-valued, we obtain from (12.11.4) the following simple analytical expressions:

$$X_1 = -\epsilon_G \frac{R_G}{Q_G}\,, \quad X_2 = \frac{R_{\max}(\epsilon_G Q_G + \epsilon_L Q_L)}{Q^2+1}\,, \quad X_3 = -\epsilon_L \frac{R_L}{Q_L} \tag{12.12.11}$$

where ϵ_G, ϵ_L are ± 1, Q_G, Q_L are given in terms of Q by Eq. (12.12.9), and either Q is given or it can be computed from Eq. (12.12.7). The choice $\epsilon_G = \epsilon_L = 1$ is made often, corresponding to capacitive X_1, X_3 and inductive X_2 [43,949].

As emphasized by Wingfield [43,949], the definition of Q as the maximum of Q_L and Q_G underestimates the total Q-factor of the network. A more appropriate definition is the sum $Q_o = Q_L + Q_G$.

An alternative set of design equations, whose input is Q_o, is obtained as follows. Given Q_o, we solve for the reference resistance R by requiring:

$$Q_o = Q_G + Q_L = \sqrt{\frac{R_G}{R} - 1} + \sqrt{\frac{R_L}{R} - 1}$$

This gives the solution for R, and hence for Q_G, Q_L:

$$\begin{aligned} R &= \frac{(R_G - R_L)^2}{(R_G + R_L)Q_o^2 - 2Q_o\sqrt{R_G R_L Q_o^2 - (R_G - R_L)^2}} \\ Q_G &= \frac{R_G Q_o - \sqrt{R_G R_L Q_o^2 - (R_G - R_L)^2}}{R_G - R_L} \\ Q_L &= \frac{R_L Q_o - \sqrt{R_G R_L Q_o^2 - (R_G - R_L)^2}}{R_L - R_G} \end{aligned} \tag{12.12.12}$$

Then, construct the Π reactances from:

$$X_1 = -\epsilon_G \frac{R_G}{Q_G}\,, \quad X_2 = R(\epsilon_G Q_G + \epsilon_L Q_L)\,, \quad X_3 = -\epsilon_L \frac{R_L}{Q_L} \tag{12.12.13}$$

The only requirement is that Q_o be greater than $Q_{\min}$. Then, it can be verified that Eqs. (12.12.12) will always result in positive values for R, Q_G, and Q_L. More simply, the value of R may be used as an input to the function `pmatch`.

Example 12.12.1: We repeat Example 12.11.1 using a Π network. Because $Z_G = 50 + 10j$ and $Z_L = 100 + 50j$, we arbitrarily choose $Z = 20 + 40j$, which satisfies $R < \min(R_G, R_L)$. The MATLAB function `pmatch` produces the solutions:

$$X_{123} = [X_1, X_2, X_3] = \text{pmatch}(Z_G, Z_L, Z) = \begin{bmatrix} 48.8304 & -71.1240 & 69.7822 \\ -35.4970 & 71.1240 & -44.7822 \\ 48.8304 & 20.5275 & -44.7822 \\ -35.4970 & -20.5275 & 69.7822 \end{bmatrix}$$

All values are in ohms and the positive ones are inductive while the negatives ones, capacitive. To see how these numbers arise, we consider the solutions of the two L-sections of Fig. 12.12.2:

$$X_{14} = \text{lmatch}(Z_G, Z, \texttt{'n'}) = \begin{bmatrix} 48.8304 & -65.2982 \\ -35.4970 & -14.7018 \end{bmatrix}$$

$$X_{35} = \text{lmatch}(Z^*, Z_L, \texttt{'r'}) = \begin{bmatrix} 69.7822 & -5.8258 \\ -44.7822 & 85.825 \end{bmatrix}$$

where X_4 and X_5 are the second columns. The four possible ways of adding the entries of X_4 and X_5 give rise to the four values of X_2. It is easily verified that each of the four solutions satisfy Eqs. (12.12.2) and (12.12.3). □

Example 12.12.2: It is desired to match a 200 ohm load to a 50 ohm source at 500 MHz. Design L-section and Π-section matching networks and compare their bandwidths.

Solution: Because $R_G < R_L$ and $X_G = 0$, only a reversed L-section will exist. Its reactances are computed from:

$$X_{12} = [X_1, X_2] = \text{lmatch}(50, 200, \texttt{'r'}) = \begin{bmatrix} 115.4701 & -86.6025 \\ -115.4701 & 86.6025 \end{bmatrix}$$

The corresponding minimum Q factor is $Q_{\min} = \sqrt{200/50 - 1} = 1.73$. Next, we design a Π section with a Q factor of 5. The required reference resistance R can be calculated from Eq. (12.12.8):

$$R = \frac{200}{5^2 + 1} = 7.6923 \text{ ohm}$$

The reactances of the Π matching section are then:

$$X_{123} = [X_1, X_2, X_3] = \text{pmatch}(50, 200, 7.6923) = \begin{bmatrix} 21.3201 & -56.5016 & 40 \\ -21.3201 & 56.5016 & -40 \\ 21.3201 & 20.4215 & -40 \\ -21.3201 & -20.4215 & 40 \end{bmatrix}$$

The Π to T transformation gives the reactances of the T-network:

$$X_{abc} = [X_a, X_b, X_c] = \text{pi2t}(X_{123}) = \begin{bmatrix} -469.0416 & 176.9861 & -250 \\ 469.0416 & -176.9861 & 250 \\ -469.0416 & -489.6805 & 250 \\ 469.0416 & 489.6805 & -250 \end{bmatrix}$$

If we increase, the Q to 15, the resulting reference resistance becomes $R = 0.885$ ohm, resulting in the reactances:

$$X_{123} = [X_1, X_2, X_3] = \text{pmatch}(50, 200, 0.885) = \begin{bmatrix} 6.7116 & -19.8671 & 13.3333 \\ -6.7116 & 19.8671 & -13.3333 \\ 6.7116 & 6.6816 & -13.3333 \\ -6.7116 & -6.6816 & 13.3333 \end{bmatrix}$$

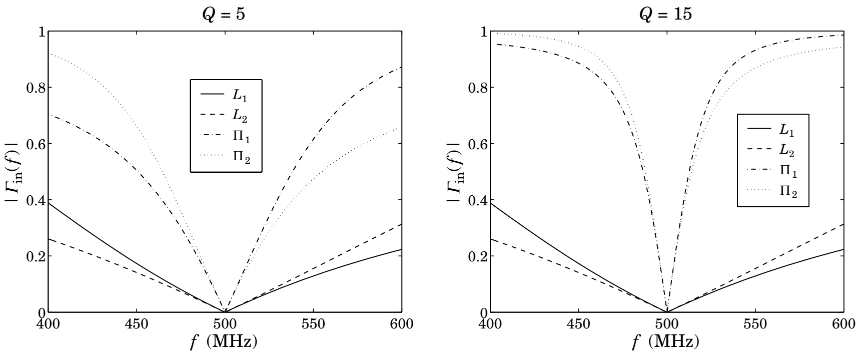


Fig. 12.12.3 Comparison of L-section and Π-section matching.

Fig. 12.12.3 shows the plot of the input reflection coefficient, that is, the quantity $\Gamma_{\rm in} = (Z_{\rm in} - Z_G^*)/(Z_{\rm in} + Z_G)$ versus frequency.

If a reactance X_i is positive, it represents an inductance with a frequency dependence of $Z_i = jX_i f/f_0$, where $f_0 = 500$ MHz is the frequency of the match. If X_i is negative, it represents a capacitance with a frequency dependence of $Z_i = jX_i f_0/f$.

The graphs display the two solutions of the L-match, but only the first two solutions of the Π match. The narrowing of the bandwidth with increasing Q is evident. □

The Π network achieves a narrower bandwidth over a single L-section network. In order to achieve a *wider* bandwidth, one may use a double L-section network [982], as shown in Fig. 12.12.4.

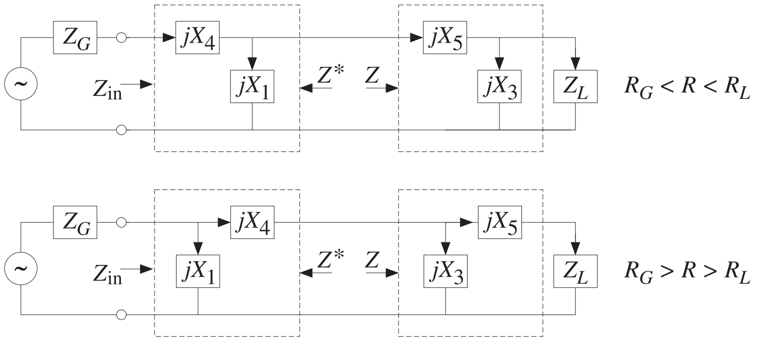


Fig. 12.12.4 Double L-section networks.

The two L-sections are either both reversed or both normal. The design is similar to Eq. (12.12.4). In particular, if $R_G < R < R_L$, we have:

$$\begin{aligned} X_{14} &= [X_1, X_4] = \text{lmatch}(Z_G, Z, \texttt{'r'}) \\ X_{35} &= [X_3, X_5] = \text{lmatch}(Z^*, Z_L, \texttt{'r'}) \end{aligned} \tag{12.12.14}$$

and if $R_G > R > R_L$:

$$\begin{aligned} X_{14} &= [X_1, X_4] = \text{lmatch}(Z_G, Z, \texttt{'n'}) \\ X_{35} &= [X_3, X_5] = \text{lmatch}(Z^*, Z_L, \texttt{'n'}) \end{aligned} \tag{12.12.15}$$

The widest bandwidth (corresponding to the smallest Q) is obtained by selecting $R = \sqrt{R_G R_L}$. For example, consider the case $R_G < R < R_L$. Then, the corresponding left and right Q factors will be:

$$Q_G = \sqrt{\frac{R}{R_G} - 1}\,, \quad Q_L = \sqrt{\frac{R_L}{R} - 1}$$

Both satisfy $Q_G < Q_{\rm min}$ and $Q_L < Q_{\rm min}$. Because we always choose Q to be the maximum of Q_G, Q_L, the optimum Q will correspond to that R that results in $Q_{\rm opt} = \min\big(\max(Q_G, Q_L)\big)$. It can be verified easily that $R_{\rm opt} = \sqrt{R_G R_L}$ and

$$Q_{\rm opt} = Q_{L,\rm opt} = Q_{G,\rm opt} = \sqrt{\frac{R_{\rm opt}}{R_G} - 1} = \sqrt{\frac{R_L}{R_{\rm opt}} - 1}$$

These results follow from the inequalities:

$$\begin{aligned} Q_G &\le Q_{\rm opt} \le Q_L\,, \quad \text{if} \quad R_G < R \le R_{\rm opt} \\ Q_L &\le Q_{\rm opt} \le Q_G\,, \quad \text{if} \quad R_{\rm opt} \le R < R_L \end{aligned}$$

Example 12.12.3: Use a double L-section to widen the bandwidth of the single L-section of Example 12.12.2.

Solution: The Q-factor of the single section is $Q_{\rm min} = \sqrt{200/500 - 1} = 1.73$. The optimum reference resistor is $R_{\rm opt} = \sqrt{50\cdot 200} = 100$ ohm and the corresponding minimized optimum $Q_{\rm opt} = 1$.

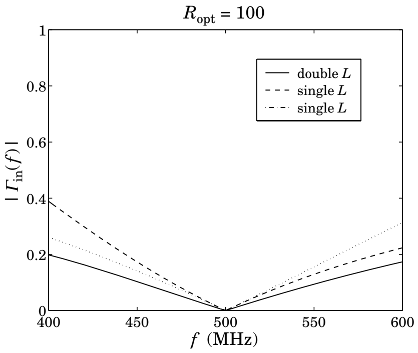


Fig. 12.12.5 Comparison of single and double L-section networks.

The reactances of the single L-section were given in Example 12.12.2. The reactances of the two sections of the double L-sections are calculated by the two calls to `lmatch`:

$$X_{14} = [X_1, X_4] = \text{lmatch}(50, 100, \texttt{'r'}) = \begin{bmatrix} 100 & -50 \\ -100 & 50 \end{bmatrix}$$

$$X_{35} = [X_3, X_5] = \text{lmatch}(100, 200, \texttt{'r'}) = \begin{bmatrix} 200 & -100 \\ -200 & 100 \end{bmatrix}$$

The corresponding input reflection coefficients are plotted in Fig. 12.12.5. As in the design of the Π network, the dual solutions of each L-section can be paired in four different ways. But, for the above optimum value of R, the four solutions have virtually identical responses. There is some widening of the bandwidth, but not by much. □

12.13 Reversed Matching Networks

The types of lossless matching networks that we considered in this chapter satisfy the property that if a network is designed to transform a load impedance Z_b into an input impedance Z_a, then the reversed (i.e., flipped left-right) network will transform the load Z_a^* into the input Z_b^*. This is illustrated in Fig. 12.13.1.

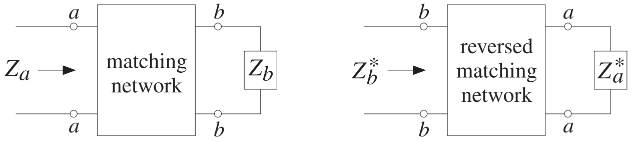


Fig. 12.13.1 Forward and reversed matching networks.

The losslessness assumption is essential. This property is satisfied only by matching networks built from segments of *lossless* transmission lines, such as stub matching or quarter-wave transformers, and by the L-, Π-, and T-section reactive networks. Some examples are shown in Fig. 12.13.2.

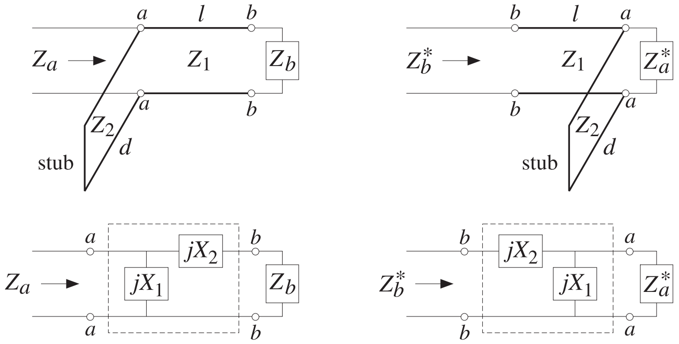


Fig. 12.13.2 Examples of reversed matching networks.

Working with admittances, we find for the stub example that the input and load admittances must be related as follows for the forward and reverse networks:

$$Y_a = Y_{\text{stub}} + Y_1 \frac{Y_b + jY_1 \tan\beta l}{Y_1 + jY_b \tan\beta l} \quad \Leftrightarrow \quad Y_b^* = Y_1 \frac{(Y_a^* + Y_{\text{stub}}) + jY_1 \tan\beta l}{Y_1 + j(Y_a^* + Y_{\text{stub}})\tan\beta l} \qquad (12.13.1)$$

where $Y_{\text{stub}} = -jY_2 \cot\beta d$ for a shorted parallel stub, and $Y_{\text{stub}} = jY_2 \tan\beta d$ for an opened one. The equivalence of the two equations in (12.13.1) is a direct consequence of the fact that Y_{stub} is purely reactive and therefore satisfies $Y_{\text{stub}}^* = -Y_{\text{stub}}$. Indeed, solving the left equation for Y_b and conjugating the answer gives:

$$Y_b = Y_1 \frac{(Y_a - Y_{\text{stub}}) - jY_1 \tan\beta l}{Y_1 - j(Y_a - Y_{\text{stub}})\tan\beta l} \quad \Rightarrow \quad Y_b^* = Y_1 \frac{(Y_a^* - Y_{\text{stub}}^*) + jY_1 \tan\beta l}{Y_1 + j(Y_a^* - Y_{\text{stub}}^*)\tan\beta l}$$

which is equivalent to the right equation (12.13.1) because $Y_{\text{stub}}^* = -Y_{\text{stub}}$. Similarly, for the L-section example we find the conditions for the forward and reversed networks:

$$Z_a = \frac{Z_1(Z_2 + Z_b)}{Z_1 + Z_2 + Z_b} \quad \Leftrightarrow \quad Z_b^* = Z_2 + \frac{Z_1 Z_a^*}{Z_1 + Z_a^*} \qquad (12.13.2)$$

where $Z_1 = jX_1$ and $Z_2 = jX_2$. The equivalence of Eqs. (12.13.2) follows again from the reactive conditions $Z_1^* = -Z_1$ and $Z_2^* = -Z_2$.

As we will see in Chap. 13, the reversing property is useful in designing the input and output matching networks of two-port networks, such as microwave amplifiers, connected to a generator and load with standardized impedance values such as $Z_0 = 50$ ohm. This is shown in Fig. 12.13.3.

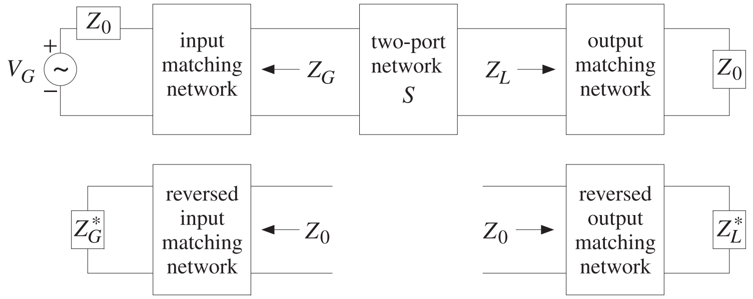


Fig. 12.13.3 Designing input and output matching networks for a two-port.

To maximize the two-port's gain or to minimize its noise figure, the two-port is required to be connected to certain optimum values of the generator and load impedances Z_G, Z_L. The output matching network must transform the actual load Z_0 into the desired value Z_L. Similarly, the input matching network must transform Z_0 into Z_G so that the two-port sees Z_G as the effective generator impedance.

In order to use the matching methods of the present chapter, it is more convenient first to design the reversed matching networks transforming a load Z_L^* (or Z_G^*) into the standardized impedance Z_0, as shown in Fig. 12.13.3. Then the designed reversed networks may be reversed to obtain the actual matching networks. Several such design examples will be presented in Chap. 13.

12.14 Problems

12.1 A one-section quarter-wavelength transformer matching a resistive load Z_L to a line Z_0 must have characteristic impedance $Z_1 = \sqrt{Z_0 Z_L}$. Show that the reflection response Γ_1 into the main line (see Fig. 12.3.1) is given as a function of frequency by:

$$\Gamma_1 = \frac{\rho(1+e^{-2j\delta})}{1+\rho^2 e^{-2j\delta}}\,, \quad \rho = \frac{\sqrt{Z_L}-\sqrt{Z_0}}{\sqrt{Z_L}+\sqrt{Z_0}}\,, \quad \delta = \frac{\pi}{2}\frac{f}{f_0}$$

where f_0 is the frequency at which the transformer length is a quarter wavelength. Show that the magnitude-squared of Γ_1 is given by:

$$|\Gamma_1|^2 = \frac{e^2\cos^2\delta}{1+e^2\cos^2\delta}\,, \quad e = \frac{2|\rho|}{1-\rho^2}$$

Show that the bandwidth (about f_0) over which the voltage standing-wave ratio on the line remains less than S is given by:

$$\sin\left(\frac{\pi}{4}\frac{\Delta f}{f_0}\right) = \frac{(S-1)(1-\rho^2)}{4|\rho|\sqrt{S}}$$

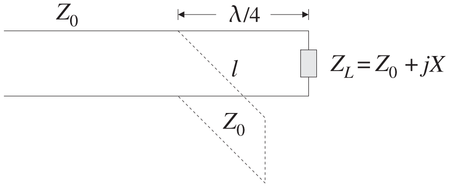
12.2 Design a one-section quarter-wavelength transformer that will match a 200-ohm load to a 50-ohm line at 100 MHz. Determine the impedance Z_1 and the bandwidth Δf over which the SWR on the line remains less than $S = 1.2$.

12.3 A transmission line with characteristic impedance $Z_0 = 100\ \Omega$ is terminated at a load impedance $Z_L = 150 + j50\ \Omega$. What percentage of the incident power is reflected back into the line?

In order to make the load reflectionless, a short-circuited stub of length l_1 and impedance also equal to Z_0 is inserted in parallel at a distance l_2 from the load. What are the smallest values of the lengths l_1 and l_2 in units of the wavelength λ that make the load reflectionless?

12.4 A loss-free line of impedance Z_0 is terminated at a load $Z_L = Z_0 + jX$, whose resistive part is matched to the line. To properly match the line, a short-circuited stub is connected across the main line at a distance of $\lambda/4$ from the load, as shown below. The stub has characteristic impedance Z_0.

Find an equation that determines the length l of the stub in order that there be no reflected waves into the main line. What is the length l (in wavelengths λ) when $X = Z_0$? When $X = Z_0/\sqrt{3}$?



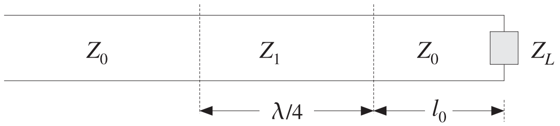
12.5 A transmission line with characteristic impedance Z_0 must be matched to a purely resistive load Z_L. A segment of length l_1 of another line of characteristic impedance Z_1 is inserted at a distance l_0 from the load, as shown in Fig. 12.7.1 (with $Z_2 = Z_0$ and $l_2 = l_0$.)

Take $Z_0 = 50$, $Z_1 = 100$, $Z_L = 80\ \Omega$ and let β_0 and β_1 be the wavenumbers within the segments l_0 and l_1. Determine the values of the quantities $\cot(\beta_1 l_1)$ and $\cot(\beta_0 l_0)$ that would guarantee matching. Show that the widest range of resistive loads Z_L that can be matched using the given values of Z_0 and Z_1 is: $12.5\ \Omega < Z_L < 200\ \Omega$.

12.6 A transmission line with resistive impedance Z_0 is terminated at a load impedance $Z_L = R + jX$. Derive an expression, in terms of Z_0, R, X, for the proportion of the incident power that is reflected back into the line.

In order to make the load reflectionless, a short-circuited stub of length l_1 and impedance Z_0 is inserted at a distance l_2 from the load. Derive expressions for the smallest values of the lengths l_1 and l_2 in terms of the wavelength λ and Z_0, R, X, that make the load reflectionless.

12.7 It is required to match a lossless transmission line Z_0 to a load Z_L. To this end, a quarter-wavelength transformer is connected at a distance l_0 from the load, as shown below. Let λ_0 and λ be the operating wavelengths of the line and the transformer segment.



Assume $Z_0 = 50\ \Omega$. Verify that the required length l_0 that will match the complex load $Z_L = 40 + 30j\ \Omega$ is $l_0 = \lambda/8$. What is the value of Z_1 in this case?

12.8 It is required to match a lossless transmission line of impedance $Z_0 = 75\ \Omega$ to the complex load $Z_L = 60 + 45j\ \Omega$. To this end, a quarter-wavelength transformer is connected at a distance l_0 from the load, as shown in the previous problem. Let λ_0 and λ be the operating wavelengths of the line and the transformer segment.

What is the required length l_0 in units of λ_0? What is the characteristic impedance Z_1 of the transformer segment?

12.9 Show that the solution of the one-section series impedance transformer shown in Fig. 12.7.3 is given by Eq. (12.7.9), provided that either of the inequalities (12.7.10) is satisfied.

12.10 Show that the solution to the double-stub tuner is given by Eq. (12.10.1) and (12.10.2).

12.11 Match load impedance $Z_L = 10 - 5j$ ohm of Example 12.8.1 to a 50-ohm line using a double-stub tuner with stub separation of $l = \lambda/16$. Show that a double-stub tuner with separation of $l = \lambda/8$ cannot match this load.

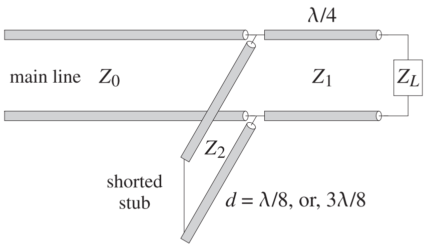
12.12 Match the antenna and feed line of Example 12.7.2 using a double stub tuner with stub separation of $l = \lambda/8$. Plot the corresponding matched reflection responses. Repeat when l is near $\lambda/2$, say, $l = 0.495\,\lambda$, and compare the resulting notch bandwidths.

12.13 Show that the load impedance of Problem 12.11 can be matched with a triple-stub tuner using shorted stubs with separations of $l_1 = l_2 = \lambda/8$, shorted stubs. Use the smallness parameter values of $e = 0.9$ and $e = 0.1$.

12.14 Match the antenna and feed line of Example 12.7.2 using a stub tuner and plot the corresponding matched reflection responses. Use shorted stubs with separations $l_1 = l_2 = \lambda/8$, and the two smallness parameters $e = 0.9$ and $e = 0.7$.

12.15 Design an L-section matching network that matches the complex load impedance $Z_L = 30 + 40j$ ohm to a 50-ohm transmission line. Verify that both a normal and a reversed L-section can be used.

12.16 It is desired to match a line with characteristic impedance Z_0 to a complex load $Z_L = R_L + jX_L$. In order to make the load reflectionless, a quarter-wavelength section of impedance Z_1 is inserted between the main line and the load, and a $\lambda/8$ or $3\lambda/8$ short-circuited stub of impedance Z_2 is inserted in parallel at the end of the line, as shown below.



a. Show that the section characteristic impedances must be chosen as:

$$Z_1 = \sqrt{Z_0 R_L}\,, \quad Z_2 = Z_0 \frac{R_L}{|X_L|}$$

Such segments are easily implemented with microstrip lines.

b. Depending on the sign of X_L, decide when one should use a $\lambda/8$ or a $3\lambda/8$ stub.

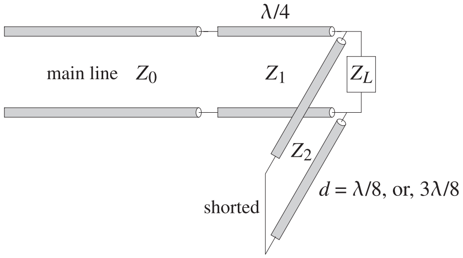
c. The above scheme works if both R_L and X_L are non-zero. What should we do if $R_L \neq 0$ and $X_L = 0$? What should we do if $R_L = 0$ and $X_L \neq 0$?

d. Repeat the above questions if an open-circuited stub is used.

12.17 A 50-ohm transmission line is terminated at the load impedance:

$$Z_L = 40 + 80j\ \Omega$$

a. In order to make the load reflectionless, a quarter-wavelength transformer section of impedance Z_1 is inserted between the line and the load, as show below, and a $\lambda/8$ or $3\lambda/8$ short-circuited stub of impedance Z_2 is inserted in parallel with the load.



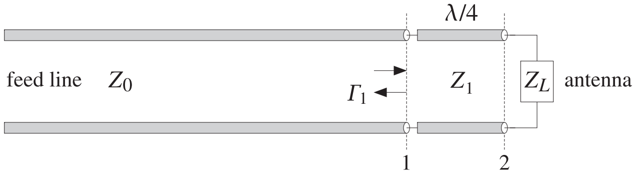
Determine the characteristic impedances Z_1 and Z_2 and whether the parallel stub should have length $\lambda/8$ or $3\lambda/8$.

b. In the general case of a shorted stub, show that the matching conditions are equivalent to the following relationship among the quantities Z_0, Z_L, Z_1, Z_2:

$$Z_L = \frac{Z_0 Z_1^2 Z_2^2 \pm j Z_2 Z_1^4}{Z_0^2 Z_2^2 + Z_1^4}$$

where Z_0, Z_1, Z_2 are assumed to be lossless. Determine which $\pm$ sign corresponds to $\lambda/8$ or $3\lambda/8$ stub length.

12.18 An FM antenna operating at a carrier frequency of $f_0 = 100$ MHz has input impedance of $Z_L = 112.5$ ohm. The antenna is to be matched to a $Z_0 = 50$ ohm feed line with a quarter-wavelength transformer inserted as shown below.



a. Determine the quarter-wavelength segment's impedance Z_1.

b. Show that the reflection response back into the feed line at the left end of the quarter-wavelength transformer is given as a function of frequency by:

$$\Gamma_1(f) = \frac{\rho(1 + e^{-2j\delta})}{1 + \rho^2 e^{-2j\delta}}\,, \quad \delta = \frac{\pi f}{2f_0}\,, \quad \rho = \frac{Z_1 - Z_0}{Z_1 + Z_0}$$

c. Plot $|\Gamma_1(f)|$ versus f in the range $0 \le f \le 200$ MHz.

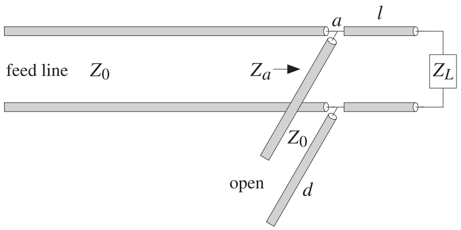
d. Using part (b), show that the bandwidth Δf_a about the carrier frequency f_0 that corresponds to a prescribed value $|\Gamma_a|^2$ of the reflection response is given by:

$$\Delta f_a = \frac{2f_0}{\pi}\,\mathrm{acos}\left(\frac{2\rho^2 - |\Gamma_a|^2(1+\rho^4)}{2\rho^2(1 - |\Gamma_a|^2)}\right)$$

e. Calculate this bandwidth for the value $|\Gamma_a| = 0.1$ and determine the left and right bandedge frequencies in MHz, and place them on the above graph of $|\Gamma_1(f)|$.

f. The FCC stipulates that FM radio stations operate within a 200 kHz bandwidth about their carrier frequency. What is the maximum value of the reflection response $|\Gamma_a|$ for such a bandwidth?

12.19 The same FM antenna is to be matched using a single-stub tuner as shown below, using an open-ended stub.



a. Determine the segment lengths d, l (in cm) assuming the segments have chacteristic impedance of $Z_0 = 50$ ohm and that the velocity factor on all the lines is 0.8.

b. Calculate and plot versus frequency the reflection response $|\Gamma_a(f)|$ into the feed line, at the terminals a shown in the figure.

13

S-Parameters

13.1 Scattering Parameters

Linear two-port (and multi-port) networks are characterized by a number of equivalent circuit parameters, such as their transfer matrix, impedance matrix, admittance matrix, and scattering matrix. Fig. 13.1.1 shows a typical two-port network.

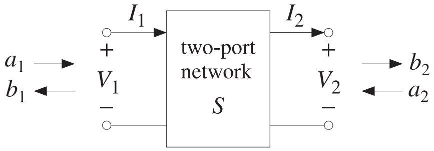


Fig. 13.1.1 Two-port network.

The *transfer matrix*, also known as the ABCD matrix, relates the voltage and current at port 1 to those at port 2, whereas the *impedance matrix* relates the two voltages V_1, V_2 to the two currents I_1, I_2:[†]

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \quad \text{(transfer matrix)}$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ -I_2 \end{bmatrix} \quad \text{(impedance matrix)} \tag{13.1.1}$$

Thus, the transfer and impedance matrices are the 2×2 matrices:

$$T = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \qquad Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{13.1.2}$$

The *admittance matrix* is simply the inverse of the impedance matrix, $Y = Z^{-1}$. The *scattering matrix* relates the *outgoing* waves b_1, b_2 to the *incoming* waves a_1, a_2 that are *incident* on the two-port:

[†] In the figure, I_2 flows out of port 2, and hence $-I_2$ flows into it. In the usual convention, both currents I_1, I_2 are taken to flow into their respective ports.

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad \text{(scattering matrix)} \qquad (13.1.3)$$

The matrix elements $S_{11}, S_{12}, S_{21}, S_{22}$ are referred to as the *scattering parameters* or the *S-parameters*. The parameters S_{11}, S_{22} have the meaning of reflection coefficients, and S_{21}, S_{12}, the meaning of transmission coefficients.

The many properties and uses of the S-parameters in applications are discussed in [956–995]. One particularly nice overview is the HP application note AN-95-1 by Anderson [971] and is available on the web [1323].

We have already seen several examples of transfer, impedance, and scattering matrices. Eq. (10.7.6) or (10.7.7) is an example of a transfer matrix and (10.8.1) is the corresponding impedance matrix. The transfer and scattering matrices of multilayer structures, Eqs. (6.6.23) and (6.6.37), are more complicated examples.

The traveling wave variables a_1, b_1 at port 1 and a_2, b_2 at port 2 are defined in terms of V_1, I_1 and V_2, I_2 and a real-valued positive reference impedance Z_0 as follows:

$$\boxed{\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} \end{aligned}} \quad \boxed{\begin{aligned} a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} \\ b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} \end{aligned}} \quad \text{(traveling waves)} \qquad (13.1.4)$$

The definitions at port 2 appear different from those at port 1, but they are really the same if expressed in terms of the incoming current $-I_2$:

$$\begin{aligned} a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{V_2 + Z_0(-I_2)}{2\sqrt{Z_0}} \\ b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{V_2 - Z_0(-I_2)}{2\sqrt{Z_0}} \end{aligned}$$

The term *traveling waves* is justified below. Eqs. (13.1.4) may be inverted to express the voltages and currents in terms of the wave variables:

$$\boxed{\begin{aligned} V_1 &= \sqrt{Z_0}\,(a_1 + b_1) \\ I_1 &= \frac{1}{\sqrt{Z_0}}\,(a_1 - b_1) \end{aligned}} \quad \boxed{\begin{aligned} V_2 &= \sqrt{Z_0}\,(a_2 + b_2) \\ I_2 &= \frac{1}{\sqrt{Z_0}}\,(b_2 - a_2) \end{aligned}} \qquad (13.1.5)$$

In practice, the reference impedance is chosen to be $Z_0 = 50$ ohm. At lower frequencies the transfer and impedance matrices are commonly used, but at microwave frequencies they become difficult to measure and therefore, the scattering matrix description is preferred.

The S-parameters can be measured by embedding the two-port network (the device-under-test, or, DUT) in a transmission line whose ends are connected to a network analyzer. Fig. 13.1.2 shows the experimental setup.

A typical network analyzer can measure S-parameters over a large frequency range, for example, the HP 8720D vector network analyzer covers the range from 50 MHz to 40 GHz. Frequency resolution is typically 1 Hz and the results can be displayed either on a Smith chart or as a conventional gain versus frequency graph.

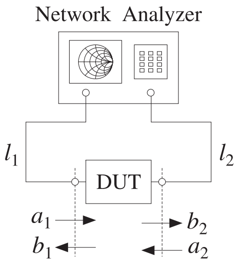


Fig. 13.1.2 Device under test connected to network analyzer.

Fig. 13.1.3 shows more details of the connection. The generator and load impedances are configured by the network analyzer. The connections can be reversed, with the generator connected to port 2 and the load to port 1.

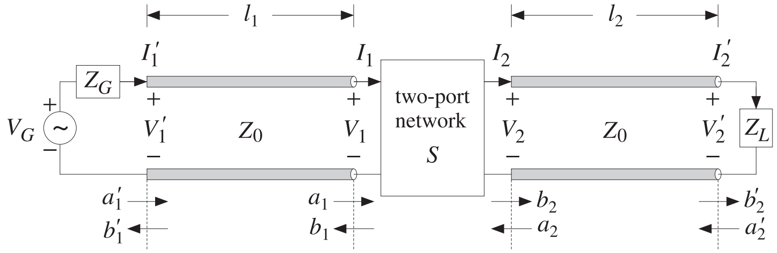


Fig. 13.1.3 Two-port network under test.

The two line segments of lengths l_1, l_2 are assumed to have characteristic impedance equal to the reference impedance Z_0. Then, the wave variables a_1, b_1 and a_2, b_2 are recognized as normalized versions of forward and backward *traveling waves*. Indeed, according to Eq. (10.7.8), we have:

$$\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{1+} \qquad & a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{2-} \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{1-} \qquad & b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{1}{\sqrt{Z_0}} V_{2+} \end{aligned} \qquad (13.1.6)$$

Thus, a_1 is essentially the incident wave at port 1 and b_1 the corresponding reflected wave. Similarly, a_2 is incident from the right onto port 2 and b_2 is the reflected wave from port 2.

The network analyzer measures the waves a_1', b_1' and a_2', b_2' at the generator and load ends of the line segments, as shown in Fig. 13.1.3. From these, the waves at the inputs of the two-port can be determined. Assuming lossless segments and using the propagation matrices (10.7.7), we have:

$$\begin{bmatrix}a_1\\ b_1\end{bmatrix}=\begin{bmatrix}e^{-j\delta_1}&0\\ 0&e^{j\delta_1}\end{bmatrix}\begin{bmatrix}a_1'\\ b_1'\end{bmatrix},\qquad \begin{bmatrix}a_2\\ b_2\end{bmatrix}=\begin{bmatrix}e^{-j\delta_2}&0\\ 0&e^{j\delta_2}\end{bmatrix}\begin{bmatrix}a_2'\\ b_2'\end{bmatrix}\tag{13.1.7}$$

where $\delta_1 = \beta l_l$ and $\delta_2 = \beta l_2$ are the phase lengths of the segments. Eqs. (13.1.7) can be rearranged into the forms:

$$\begin{bmatrix}b_1\\ b_2\end{bmatrix}=D\begin{bmatrix}b_1'\\ b_2'\end{bmatrix},\qquad \begin{bmatrix}a_1'\\ a_2'\end{bmatrix}=D\begin{bmatrix}a_1\\ a_2\end{bmatrix},\qquad D=\begin{bmatrix}e^{j\delta_1}&0\\ 0&e^{j\delta_2}\end{bmatrix}$$

The network analyzer measures the corresponding S-parameters of the primed variables, that is,

$$\begin{bmatrix}b_1'\\ b_2'\end{bmatrix}=\begin{bmatrix}S_{11}'&S_{12}'\\ S_{21}'&S_{22}'\end{bmatrix}\begin{bmatrix}a_1'\\ a_2'\end{bmatrix},\qquad S'=\begin{bmatrix}S_{11}'&S_{12}'\\ S_{21}'&S_{22}'\end{bmatrix}\quad\text{(measured S-matrix)}\tag{13.1.8}$$

The S-matrix of the two-port can be obtained then from:

$$\begin{bmatrix}b_1\\ b_2\end{bmatrix}=D\begin{bmatrix}b_1'\\ b_2'\end{bmatrix}=DS'\begin{bmatrix}a_1'\\ a_2'\end{bmatrix}=DS'D\begin{bmatrix}a_1\\ a_2\end{bmatrix}\quad\Rightarrow\quad S=DS'D$$

or, more explicitly,

$$\begin{aligned}\begin{bmatrix}S_{11}&S_{12}\\ S_{21}&S_{22}\end{bmatrix}&=\begin{bmatrix}e^{j\delta_1}&0\\ 0&e^{j\delta_2}\end{bmatrix}\begin{bmatrix}S_{11}'&S_{12}'\\ S_{21}'&S_{22}'\end{bmatrix}\begin{bmatrix}e^{j\delta_1}&0\\ 0&e^{j\delta_2}\end{bmatrix}\\ &=\begin{bmatrix}S_{11}'e^{2j\delta_1}&S_{12}'e^{j(\delta_1+\delta_2)}\\ S_{21}'e^{j(\delta_1+\delta_2)}&S_{22}'e^{2j\delta_2}\end{bmatrix}\end{aligned}\tag{13.1.9}$$

Thus, changing the points along the transmission lines at which the S-parameters are measured introduces only phase changes in the parameters.

Without loss of generality, we may replace the extended circuit of Fig. 13.1.3 with the one shown in Fig. 13.1.4 with the understanding that either we are using the extended two-port parameters S', or, equivalently, the generator and segment l_1 have been replaced by their Thévenin equivalents, and the load impedance has been replaced by its propagated version to distance l_2.

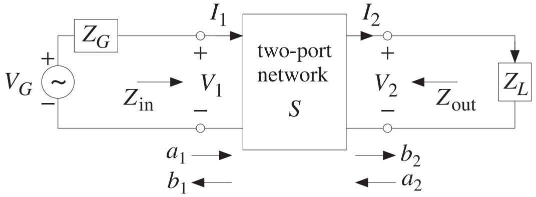


Fig. 13.1.4 Two-port network connected to generator and load.

The actual measurements of the S-parameters are made by connecting to a matched load, $Z_L = Z_0$. Then, there will be no reflected waves from the load, $a_2 = 0$, and the S-matrix equations will give:

$$b_1 = S_{11}a_1 + S_{12}a_2 = S_{11}a_1 \quad\Rightarrow\quad S_{11} = \frac{b_1}{a_1}\bigg|_{Z_L=Z_0} = \text{reflection coefficient}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 = S_{21}a_1 \quad\Rightarrow\quad S_{21} = \frac{b_2}{a_1}\bigg|_{Z_L=Z_0} = \text{transmission coefficient}$$

Reversing the roles of the generator and load, one can measure in the same way the parameters S_{12} and S_{22}.

13.2 Power Flow

Power flow into and out of the two-port is expressed very simply in terms of the traveling wave amplitudes. Using the inverse relationships (13.1.5), we find:

$$\begin{aligned}\frac{1}{2}\text{Re}[V_1^*I_1] &= \frac{1}{2}|a_1|^2-\frac{1}{2}|b_1|^2\\ -\frac{1}{2}\text{Re}[V_2^*I_2] &= \frac{1}{2}|a_2|^2-\frac{1}{2}|b_2|^2\end{aligned}\tag{13.2.1}$$

The left-hand sides represent the power flow *into* ports 1 and 2. The right-hand sides represent the difference between the power incident on a port and the power reflected from it. The quantity $\text{Re}[V_2^*I_2]/2$ represents the power transferred to the load.

Another way of phrasing these is to say that part of the incident power on a port gets reflected and part enters the port:

$$\begin{aligned}\frac{1}{2}|a_1|^2 &= \frac{1}{2}|b_1|^2+\frac{1}{2}\text{Re}[V_1^*I_1]\\ \frac{1}{2}|a_2|^2 &= \frac{1}{2}|b_2|^2+\frac{1}{2}\text{Re}[V_2^*(-I_2)]\end{aligned}\tag{13.2.2}$$

One of the reasons for normalizing the traveling wave amplitudes by $\sqrt{Z_0}$ in the definitions (13.1.4) was precisely this simple way of expressing the incident and reflected powers from a port.

If the two-port is lossy, the power lost in it will be the difference between the power entering port 1 and the power leaving port 2, that is,

$$P_{\text{loss}} = \frac{1}{2}\text{Re}[V_1^*I_1]-\frac{1}{2}\text{Re}[V_2^*I_2]=\frac{1}{2}|a_1|^2+\frac{1}{2}|a_2|^2-\frac{1}{2}|b_1|^2-\frac{1}{2}|b_2|^2$$

Noting that $\mathbf{a}^\dagger\mathbf{a} = |a_1|^2+|a_2|^2$ and $\mathbf{b}^\dagger\mathbf{b} = |b_1|^2+|b_2|^2$, and writing $\mathbf{b}^\dagger\mathbf{b} = \mathbf{a}^\dagger S^\dagger S\mathbf{a}$, we may express this relationship in terms of the scattering matrix:

$$P_{\text{loss}} = \frac{1}{2}\mathbf{a}^\dagger\mathbf{a}-\frac{1}{2}\mathbf{b}^\dagger\mathbf{b}=\frac{1}{2}\mathbf{a}^\dagger\mathbf{a}-\frac{1}{2}\mathbf{a}^\dagger S^\dagger S\mathbf{a}=\frac{1}{2}\mathbf{a}^\dagger(I-S^\dagger S)\mathbf{a}\tag{13.2.3}$$

For a lossy two-port, the power loss is positive, which implies that the matrix $I - S^\dagger S$ must be positive definite. If the two-port is lossless, $P_{\text{loss}} = 0$, the S-matrix will be *unitary*, that is, $S^\dagger S = I$.

We already saw examples of such unitary scattering matrices in the cases of the equal travel-time multilayer dielectric structures and their equivalent quarter wavelength multisection transformers.

13.3 Parameter Conversions

It is straightforward to derive the relationships that allow one to pass from one parameter set to another. For example, starting with the transfer matrix, we have:

$$\begin{aligned} V_1 &= AV_2 + BI_2 \\ I_1 &= CV_2 + DI_2 \end{aligned} \quad\Rightarrow\quad \begin{aligned} V_1 &= A\Big(\frac{1}{C}I_1 - \frac{D}{C}I_2\Big) + BI_2 = \frac{A}{C}I_1 - \frac{AD-BC}{C}I_2 \\ V_2 &= \frac{1}{C}I_1 - \frac{D}{C}I_2 \end{aligned}$$

The coefficients of I_1, I_2 are the impedance matrix elements. The steps are reversible, and we summarize the final relationships below:

$$\begin{aligned} Z &= \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = \frac{1}{C}\begin{bmatrix} A & AD-BC \\ 1 & D \end{bmatrix} \\ T &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{Z_{21}}\begin{bmatrix} Z_{11} & Z_{11}Z_{22} - Z_{12}Z_{21} \\ 1 & Z_{22} \end{bmatrix} \end{aligned} \tag{13.3.1}$$

We note the determinants $\det(T) = AD - BC$ and $\det(Z) = Z_{11}Z_{22} - Z_{12}Z_{21}$. The relationship between the scattering and impedance matrices is also straightforward to derive. We define the 2×1 vectors:

$$\boldsymbol{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad \boldsymbol{I} = \begin{bmatrix} I_1 \\ -I_2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{13.3.2}$$

Then, the definitions (13.1.4) can be written compactly as:

$$\begin{aligned} \mathbf{a} &= \frac{1}{2\sqrt{Z_0}}(\boldsymbol{V} + Z_0\boldsymbol{I}) = \frac{1}{2\sqrt{Z_0}}(Z + Z_0 I)\boldsymbol{I} \\ \mathbf{b} &= \frac{1}{2\sqrt{Z_0}}(\boldsymbol{V} - Z_0\boldsymbol{I}) = \frac{1}{2\sqrt{Z_0}}(Z - Z_0 I)\boldsymbol{I} \end{aligned} \tag{13.3.3}$$

where we used the impedance matrix relationship $\boldsymbol{V} = Z\boldsymbol{I}$ and defined the 2×2 unit matrix I. It follows then,

$$\frac{1}{2\sqrt{Z_0}}\boldsymbol{I} = (Z + Z_0 I)^{-1}\mathbf{a} \quad\Rightarrow\quad \mathbf{b} = \frac{1}{2\sqrt{Z_0}}(Z - Z_0 I)\boldsymbol{I} = (Z - Z_0 I)(Z + Z_0 I)^{-1}\mathbf{a}$$

Thus, the scattering matrix S will be related to the impedance matrix Z by

$$\boxed{S = (Z - Z_0 I)(Z + Z_0 I)^{-1}} \quad\Leftrightarrow\quad \boxed{Z = (I - S)^{-1}(I + S)Z_0} \tag{13.3.4}$$

Explicitly, we have:

$$\begin{aligned} S &= \begin{bmatrix} Z_{11} - Z_0 & Z_{12} \\ Z_{21} & Z_{22} - Z_0 \end{bmatrix}\begin{bmatrix} Z_{11} + Z_0 & Z_{12} \\ Z_{21} & Z_{22} + Z_0 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} Z_{11} - Z_0 & Z_{12} \\ Z_{21} & Z_{22} - Z_0 \end{bmatrix}\frac{1}{D_z}\begin{bmatrix} Z_{22} + Z_0 & -Z_{12} \\ -Z_{21} & Z_{11} + Z_0 \end{bmatrix} \end{aligned}$$

where $D_z = \det(Z + Z_0 I) = (Z_{11} + Z_0)(Z_{22} + Z_0) - Z_{12}Z_{21}$. Multiplying the matrix factors, we obtain:

$$S = \frac{1}{D_z}\begin{bmatrix} (Z_{11} - Z_0)(Z_{22} + Z_0) - Z_{12}Z_{21} & 2Z_{12}Z_0 \\ 2Z_{21}Z_0 & (Z_{11} + Z_0)(Z_{22} - Z_0) - Z_{12}Z_{21} \end{bmatrix} \tag{13.3.5}$$

Similarly, the inverse relationship gives:

$$Z = \frac{Z_0}{D_s}\begin{bmatrix} (1 + S_{11})(1 - S_{22}) + S_{12}S_{21} & 2S_{12} \\ 2S_{21} & (1 - S_{11})(1 + S_{22}) + S_{12}S_{21} \end{bmatrix} \tag{13.3.6}$$

where $D_s = \det(I - S) = (1 - S_{11})(1 - S_{22}) - S_{12}S_{21}$. Expressing the impedance parameters in terms of the transfer matrix parameters, we also find:

$$S = \frac{1}{D_a}\begin{bmatrix} A + \dfrac{B}{Z_0} - CZ_0 - D & 2(AD - BC) \\ 2 & -A + \dfrac{B}{Z_0} - CZ_0 + D \end{bmatrix} \tag{13.3.7}$$

where $D_a = A + \dfrac{B}{Z_0} + CZ_0 + D$.

13.4 Input and Output Reflection Coefficients

When the two-port is connected to a generator and load as in Fig. 13.1.4, the impedance and scattering matrix equations take the simpler forms:

$$\boxed{\begin{aligned} V_1 &= Z_{\text{in}} I_1 \\ V_2 &= Z_L I_2 \end{aligned}} \quad\Leftrightarrow\quad \boxed{\begin{aligned} b_1 &= \Gamma_{\text{in}} a_1 \\ a_2 &= \Gamma_L b_2 \end{aligned}} \tag{13.4.1}$$

where Z_{in} is the input impedance at port 1, and $\Gamma_{\text{in}}, \Gamma_L$ are the reflection coefficients at port 1 and at the load:

$$\Gamma_{\text{in}} = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0}, \quad \Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{13.4.2}$$

The input impedance and input reflection coefficient can be expressed in terms of the Z- and S-parameters, as follows:

$$\boxed{Z_{\text{in}} = Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22} + Z_L}} \quad\Leftrightarrow\quad \boxed{\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}} \tag{13.4.3}$$

The equivalence of these two expressions can be shown by using the parameter conversion formulas of Eqs. (13.3.5) and (13.3.6), or they can be shown indirectly, as follows. Starting with $V_2 = Z_L I_2$ and using the second impedance matrix equation, we can solve for I_2 in terms of I_1:

$$V_2 = Z_{21}I_1 - Z_{22}I_2 = Z_L I_2 \quad \Rightarrow \quad I_2 = \frac{Z_{21}}{Z_{22} + Z_L}I_1 \tag{13.4.4}$$

Then, the first impedance matrix equation implies:

$$V_1 = Z_{11}I_1 - Z_{12}I_2 = \left(Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22} + Z_L}\right)I_1 = Z_{\text{in}}I_1$$

Starting again with $V_2 = Z_L I_2$ we find for the traveling waves at port 2:

$$\begin{aligned} a_2 &= \frac{V_2 - Z_0 I_2}{2\sqrt{Z_0}} = \frac{Z_L - Z_0}{2\sqrt{Z_0}}I_2 \\ b_2 &= \frac{V_2 + Z_0 I_2}{2\sqrt{Z_0}} = \frac{Z_L + Z_0}{2\sqrt{Z_0}}I_2 \end{aligned} \quad \Rightarrow \quad a_2 = \frac{Z_L - Z_0}{Z_L + Z_0}b_2 = \Gamma_L b_2$$

Using $V_1 = Z_{\text{in}}I_1$, a similar argument implies for the waves at port 1:

$$\begin{aligned} a_1 &= \frac{V_1 + Z_0 I_1}{2\sqrt{Z_0}} = \frac{Z_{\text{in}} + Z_0}{2\sqrt{Z_0}}I_1 \\ b_1 &= \frac{V_1 - Z_0 I_1}{2\sqrt{Z_0}} = \frac{Z_{\text{in}} - Z_0}{2\sqrt{Z_0}}I_1 \end{aligned} \quad \Rightarrow \quad b_1 = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0}a_1 = \Gamma_{\text{in}}a_1$$

It follows then from the scattering matrix equations that:

$$b_2 = S_{21}a_1 + S_{22}a_2 = S_{22}a_1 + S_{22}\Gamma_L b_2 \quad \Rightarrow \quad \boxed{b_2 = \frac{S_{21}}{1 - S_{22}\Gamma_L}a_1} \tag{13.4.5}$$

which implies for b_1:

$$b_1 = S_{11}a_1 + S_{12}a_2 = S_{11}a_1 + S_{12}\Gamma_L b_2 = \left(S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}\right)a_1 = \Gamma_{\text{in}}a_1$$

Reversing the roles of generator and load, we obtain the impedance and reflection coefficients from the output side of the two-port:

$$\boxed{Z_{\text{out}} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11} + Z_G}} \quad \Leftrightarrow \quad \boxed{\Gamma_{\text{out}} = S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1 - S_{11}\Gamma_G}} \tag{13.4.6}$$

where

$$\Gamma_{\text{out}} = \frac{Z_{\text{out}} - Z_0}{Z_{\text{out}} + Z_0}, \quad \Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0} \tag{13.4.7}$$

The input and output impedances allow one to replace the original two-port circuit of Fig. 13.1.4 by simpler equivalent circuits. For example, the two-port and the load can be replaced by the input impedance Z_{in} connected at port 1, as shown in Fig. 13.4.1.

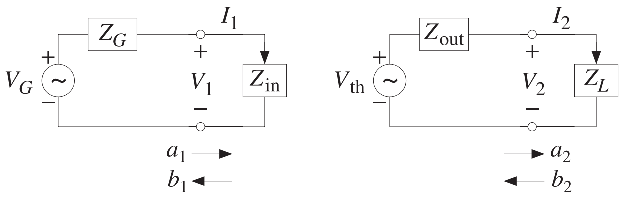


Fig. 13.4.1 Input and output equivalent circuits.

Similarly, the generator and the two-port can be replaced by a Thévenin equivalent circuit connected at port 2. By determining the open-circuit voltage and short-circuit current at port 2, we find the corresponding Thévenin parameters in terms of the impedance parameters:

$$\boxed{V_{\text{th}} = \frac{Z_{21}V_G}{Z_{11} + Z_G}, \quad Z_{\text{th}} = Z_{\text{out}} = Z_{22} - \frac{Z_{12}Z_{21}}{Z_{11} + Z_G}} \tag{13.4.8}$$

13.5 Stability Circles

In discussing the stability conditions of a two-port in terms of S-parameters, the following definitions of constants are often used:

$$\begin{aligned} \Delta &= \det(S) = S_{11}S_{22} - S_{12}S_{21} \\ K &= \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \qquad \text{(Rollett stability factor)} \\ \mu_1 &= \frac{1 - |S_{11}|^2}{|S_{22} - \Delta S_{11}^*| + |S_{12}S_{21}|} \qquad \text{(Edwards-Sinsky stability parameter)} \\ \mu_2 &= \frac{1 - |S_{22}|^2}{|S_{11} - \Delta S_{22}^*| + |S_{12}S_{21}|} \\ B_1 &= 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \\ B_2 &= 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \\ C_1 &= S_{11} - \Delta S_{22}^*, \quad D_1 = |S_{11}|^2 - |\Delta|^2 \\ C_2 &= S_{22} - \Delta S_{11}^*, \quad D_2 = |S_{22}|^2 - |\Delta|^2 \end{aligned} \tag{13.5.1}$$

The quantity K is the Rollett stability factor [967], and μ_1, μ_2, the Edwards-Sinsky stability parameters [970]. The following identities hold among these constants:

$$\begin{aligned} B_1^2 - 4|C_1|^2 &= B_2^2 - 4|C_2|^2 = 4|S_{12}S_{21}|^2(K^2 - 1) \\ |C_1|^2 &= |S_{12}S_{21}|^2 + (1 - |S_{22}|^2)D_1 \\ |C_2|^2 &= |S_{12}S_{21}|^2 + (1 - |S_{11}|^2)D_2 \end{aligned} \tag{13.5.2}$$

For example, noting that $S_{12}S_{21} = S_{11}S_{22} - \Delta$, the last of Eqs. (13.5.2) is a direct consequence of the identity:

$$|A - BC|^2 - |B - AC^*|^2 = \big(1 - |C|^2\big)\big(|A|^2 - |B|^2\big) \tag{13.5.3}$$

We define also the following parameters, which will be recognized as the centers and radii of the source and load stability circles:

$$c_G = \frac{C_1^*}{D_1}\,, \quad r_G = \frac{|S_{12}S_{21}|}{|D_1|} \qquad \text{(source stability circle)} \tag{13.5.4}$$

$$c_L = \frac{C_2^*}{D_2}\,, \quad r_L = \frac{|S_{12}S_{21}|}{|D_2|} \qquad \text{(load stability circle)} \tag{13.5.5}$$

They satisfy the following relationships, which are consequences of the last two of Eqs. (13.5.2) and the definitions (13.5.4) and (13.5.5):

$$\begin{aligned} 1 - |S_{11}|^2 &= \big(|c_L|^2 - r_L^2\big)D_2 \\ 1 - |S_{22}|^2 &= \big(|c_G|^2 - r_G^2\big)D_1 \end{aligned} \tag{13.5.6}$$

We note also that using Eqs. (13.5.6), the stability parameters μ_1, μ_2 can be written as:

$$\begin{aligned} \mu_1 &= \big(|c_L| - r_L\big)\,\mathrm{sign}(D_2) \\ \mu_2 &= \big(|c_G| - r_G\big)\,\mathrm{sign}(D_1) \end{aligned} \tag{13.5.7}$$

For example, we have:

$$\mu_1 = \frac{1 - |S_{11}|^2}{|C_2| + |S_{12}S_{21}|} = \frac{D_2\big(|c_L|^2 - r_L^2\big)}{|D_2||c_L| + |D_2|r_L} = \frac{D_2\big(|c_L|^2 - r_L^2\big)}{|D_2|\big(|c_L| + r_L\big)} = \frac{D_2}{|D_2|}\big(|c_L| - r_L\big)$$

We finally note that the input and output reflection coefficients can be written in the alternative forms:

$$\begin{aligned} \Gamma_{\rm in} &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \\ \Gamma_{\rm out} &= S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1 - S_{22}\Gamma_G} = \frac{S_{22} - \Delta\Gamma_G}{1 - S_{11}\Gamma_G} \end{aligned} \tag{13.5.8}$$

Next, we discuss the stability conditions. The two-port is *unconditionally stable* if any generator and load impedances with positive resistive parts R_G, R_L, will always lead to input and output impedances with positive resistive parts $R_{\rm in}, R_{\rm out}$.

Equivalently, unconditional stability requires that any load and generator with $|\Gamma_L| < 1$ and $|\Gamma_G| < 1$ will result into $|\Gamma_{\rm in}| < 1$ and $|\Gamma_{\rm out}| < 1$.

The two-port is termed *potentially or conditionally unstable* if there are $|\Gamma_L| < 1$ and $|\Gamma_G| < 1$ resulting into $|\Gamma_{\rm in}| \ge 1$ and/or $|\Gamma_{\rm out}| \ge 1$.

The *load stability region* is the set of all Γ_L that result into $|\Gamma_{\rm in}| < 1$, and the *source stability region*, the set of all Γ_G that result into $|\Gamma_{\rm out}| < 1$.

In the unconditionally stable case, the load and source stability regions contain the entire unit-circles $|\Gamma_L| < 1$ or $|\Gamma_G| < 1$. However, in the potentially unstable case, only portions of the unit-circles may lie within the stability regions and such Γ_G, Γ_L will lead to a stable input and output impedances.

The connection of the stability regions to the stability circles is brought about by the following identities, which can be proved easily using Eqs. (13.5.1)–(13.5.8):

$$\begin{aligned} 1 - |\Gamma_{\rm in}|^2 &= \frac{|\Gamma_L - c_L|^2 - r_L^2}{|1 - S_{22}\Gamma_L|^2}\,D_2 \\ 1 - |\Gamma_{\rm out}|^2 &= \frac{|\Gamma_G - c_G|^2 - r_G^2}{|1 - S_{11}\Gamma_G|^2}\,D_1 \end{aligned} \tag{13.5.9}$$

For example, the first can be shown starting with Eq. (13.5.8) and using the definitions (13.5.5) and the relationship (13.5.6):

$$\begin{aligned} 1 - |\Gamma_{\rm in}|^2 = 1 - \left|\frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L}\right|^2 &= \frac{|S_{11} - \Delta\Gamma_L|^2 - |1 - S_{22}\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{\big(|S_{22}|^2 - |\Delta|^2\big)|\Gamma_L|^2 - (S_{22} - \Delta S_{11}^*)\Gamma_L - (S_{22}^* - \Delta^* S_{11})\Gamma_L^* + 1 - |S_{11}|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{D_2|\Gamma_L|^2 - C_2\Gamma_L - C_2^*\Gamma_L^* + 1 - |S_{11}|^2}{|1 - S_{22}\Gamma_L|^2} \\ &= \frac{D_2\big(|\Gamma_L|^2 - c_L^*\Gamma_L - c_L^*\Gamma_L^* + |c_L|^2 - r_L^2\big)}{|1 - S_{22}\Gamma_L|^2} = \frac{D_2\big(|\Gamma_L - c_L|^2 - r_L^2\big)}{|1 - S_{22}\Gamma_L|^2} \end{aligned}$$

It follows from Eq. (13.5.9) that the load stability region is defined by the conditions:

$$1 - |\Gamma_{\rm in}|^2 > 0 \quad \Leftrightarrow \quad \big(|\Gamma_L - c_L|^2 - r_L^2\big)D_2 > 0$$

Depending on the sign of D_2, these are equivalent to the outside or the inside of the load stability circle of center c_L and radius r_L:

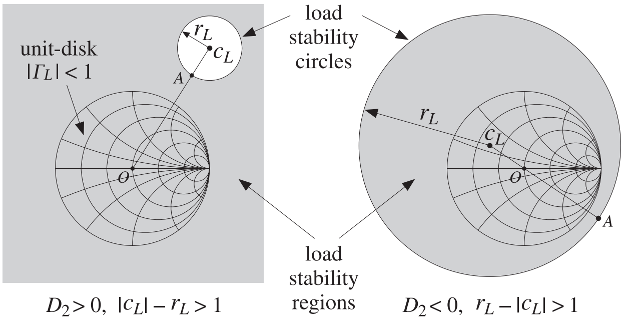
$$\boxed{\begin{aligned} |\Gamma_L - c_L| > r_L\,, &\quad \text{if } D_2 > 0 \\ |\Gamma_L - c_L| < r_L\,, &\quad \text{if } D_2 < 0 \end{aligned}} \qquad \text{(load stability region)} \tag{13.5.10}$$

The boundary of the circle $|\Gamma_L - c_L| = r_L$ corresponds to $|\Gamma_{\rm in}| = 1$. The complement of these regions corresponds to the unstable region with $|\Gamma_{\rm in}| > 1$. Similarly, we find for the source stability region:

$$\boxed{\begin{aligned} |\Gamma_G - c_G| > r_G\,, &\quad \text{if } D_1 > 0 \\ |\Gamma_G - c_G| < r_G\,, &\quad \text{if } D_1 < 0 \end{aligned}} \qquad \text{(source stability region)} \tag{13.5.11}$$

In order to have unconditional stability, the stability regions must contain the unit-circle in its entirety. If $D_2 > 0$, the unit-circle and load stability circle must not overlap at all, as shown in Fig. 13.5.1. Geometrically, the distance between the points O and A in the figure is $(OA) = |c_L| - r_L$. The non-overlapping of the circles requires the condition $(OA) > 1$, or, $|c_L| - r_L > 1$.

If $D_2 < 0$, the stability region is the inside of the stability circle, and therefore, the unit-circle must lie within that circle. This requires that $(OA) = r_L - |c_L| > 1$, as shown in Fig. 13.5.1.



$D_2 > 0,\ |c_L| - r_L > 1$ $D_2 < 0,\ r_L - |c_L| > 1$

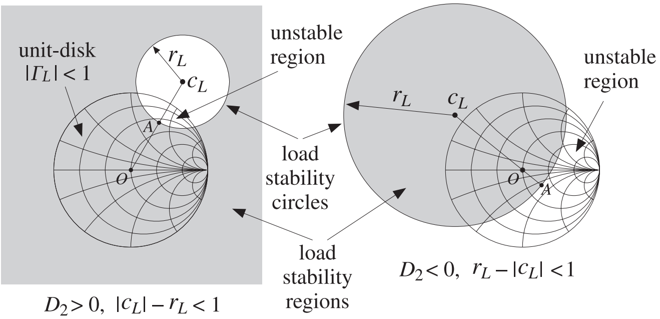
Fig. 13.5.1 Load stability regions in the unconditionally stable case.

These two conditions can be combined into $\text{sign}(D_2)\big(|c_L| - r_L\big) > 1$. But, that is equivalent to $\mu_1 > 1$ according to Eq. (13.5.7). Geometrically, the parameter μ_1 represents the distance (OA). Thus, the condition for the unconditional stability of the input is equivalent to:

$$\boxed{\mu_1 > 1} \qquad \text{(unconditional stability condition)} \tag{13.5.12}$$

It has been shown by Edwards and Sinsky [970] that this single condition (or, alternatively, the single condition $\mu_2 > 1$) is necessary and sufficient for the unconditional stability of both the input and output impedances of the two-port. Clearly, the source stability regions will be similar to those of Fig. 13.5.1.

If the stability condition is not satisfied, that is, $\mu_1 < 1$, then only that portion of the unit-circle that lies within the stability region will be stable and will lead to stable input and output impedances. Fig. 13.5.2 illustrates such a potentially unstable case.



$D_2 > 0,\ |c_L| - r_L < 1$ $D_2 < 0,\ r_L - |c_L| < 1$

Fig. 13.5.2 Load stability regions in potentially unstable case.

If $D_2 > 0$, then $\mu_1 < 1$ is equivalent to $|c_L| - r_L < 1$, and if $D_2 < 0$, it is equivalent to $r_L - |c_L| < 1$. In either case, the unit-circle is partially overlapping with the stability circle, as shown in Fig. 13.5.2. The portion of the unit-circle that does not lie within the stability region will correspond to an unstable Z_{in}.

There exist several other unconditional stability criteria that are equivalent to the single criterion $\mu_1 > 1$. They all require that the Rollett stability factor K be greater than unity, $K > 1$, as well as one other condition. Any one of the following criteria are necessary and sufficient for unconditional stability [968]:

$$\begin{array}{lll} K > 1 & \text{and} & |\Delta| < 1 \\ K > 1 & \text{and} & B_1 > 0 \\ K > 1 & \text{and} & B_2 > 0 \\ K > 1 & \text{and} & |S_{12}S_{21}| < 1 - |S_{11}|^2 \\ K > 1 & \text{and} & |S_{12}S_{21}| < 1 - |S_{22}|^2 \end{array} \qquad \text{(stability conditions)} \tag{13.5.13}$$

Their equivalence to $\mu_1 > 1$ has been shown in [970]. In particular, it follows from the last two conditions that unconditional stability requires $|S_{11}| < 1$ and $|S_{22}| < 1$. These are necessary but not sufficient for stability.

A very common circumstance in practice is to have a potentially unstable two-port, but with $|S_{11}| < 1$ and $|S_{22}| < 1$. In such cases, Eq. (13.5.6) implies $D_2\big(|c_L|^2 - r_L^2\big) > 0$, and the lack of stability requires $\mu_1 = \text{sign}(D_2)\big(|c_L|^2 - r_L^2\big) < 1$.

Therefore, if $D_2 > 0$, then we must have $|c_L|^2 - r_L^2 > 0$ and $|c_L| - r_L < 1$, which combine into the inequality $r_L < |c_L| < r_L + 1$. This is depicted in the left picture of Fig. 13.5.2. The geometrical distance $(OA) = |c_L| - r_L$ satisfies $0 < (OA) < 1$, so that stability circle partially overlaps with the unit-circle but does not enclose its center.

On the other hand, if $D_2 < 0$, the two conditions require $|c_L|^2 - r_L^2 < 0$ and $r_L - |c_L| < 1$, which imply $|c_L| < r_L < |c_L| + 1$. This is depicted in the right Fig. 13.5.2. The geometrical distance $(OA) = r_L - |c_L|$ again satisfies $0 < (OA) < 1$, but now the center of the unit-circle lies within the stability circle, which is also the stability region.

We have written a number of MATLAB functions that facilitate working with S-parameters. They are described in detail later on:

`smat`	reshape S-parameters into S-matrix
`sparam`	calculate stability parameters
`sgain`	calculate transducer, available, operating, and unilateral power gains
`smatch`	calculate simultaneous conjugate match for generator and load
`gin,gout`	calculate input and output reflection coefficients
`smith`	draw a basic Smith chart
`smithcir`	draw a stability or gain circle on Smith chart
`sgcirc`	determine stability and gain circles
`nfcirc`	determine noise figure circles
`nfig`	calculate noise figure

The MATLAB function `sparam` calculates the stability parameters $\mu_1, K, |\Delta|, B_1, B_2$, as well as the parameters C_1, C_2, D_1, D_2. It has usage:

```
[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);        % stability parameters
```

The function `sgcirc` calculates the centers and radii of the source and load stability circles. It also calculates gain circles to be discussed later on. Its usage is:

```
[cL,rL] = sgcirc(S,'l');          % load or Zin stability circle
[cG,rG] = sgcirc(S,'s');          % source or Zout stability circle
```

The MATLAB function `smith` draws a basic Smith chart, and the function `smithcir` draws the stability circles:

```
smith(n);                         % draw four basic types of Smith charts, n = 1,2,3,4
smith;                            % default Smith chart corresponding to n = 3

smithcir(c,r,max,width);          % draw circle of center c and radius r
smithcir(c,r,max);                % equivalent to linewidth width=1
smithcir(c,r);                    % draw full circle with linewidth width=1
```

The parameter `max` controls the portion of the stability circle that is visible outside the Smith chart. For example, `max` $= 1.1$ will display only that portion of the circle that has $|\Gamma| < 1.1$.

Example 13.5.1: The Hewlett-Packard AT-41511 NPN bipolar transistor has the following S-parameters at 1 GHz and 2 GHz [1324]:

$$\begin{array}{llll} S_{11} = 0.48\angle{-149^o}\,, & S_{21} = 5.189\angle 89^o\,, & S_{12} = 0.073\angle 43^o\,, & S_{22} = 0.49\angle{-39^o} \\ S_{11} = 0.46\angle 162^o\,, & S_{21} = 2.774\angle 59^o\,, & S_{12} = 0.103\angle 45^o\,, & S_{22} = 0.42\angle{-47^o} \end{array}$$

Determine the stability parameters, stability circles, and stability regions.

Solution: The transistor is potentially unstable at 1 GHz, but unconditionally stable at 2 GHz. The source and load stability circles at 1 GHz are shown in Fig. 13.5.3.

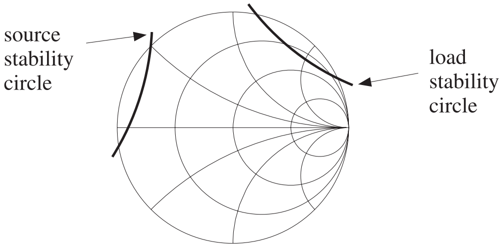


Fig. 13.5.3 Load and source stability circles at 1 GHz.

The MATLAB code used to generate this graph was:

```
S = smat([0.48 -149 5.189 89 0.073 43 0.49 -39]);    % form S-matrix

[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);              % stability parameters

[cL,rL] = sgcirc(S,'l');                             % stability circles
[cG,rG] = sgcirc(S,'s');

smith;                                               % draw basic Smith chart
smithcir(cL, rL, 1.1, 1.5);                          % draw stability circles
smithcir(cG, rG, 1.1, 1.5);
```

The computed stability parameters at 1 GHz were:

$$[K,\mu_1,|\Delta|,B_1,B_2,D_1,D_2] = [0.781,\ 0.847,\ 0.250,\ 0.928,\ 0.947,\ 0.168,\ 0.178]$$

The transistor is potentially unstable because $K < 1$ even though $|\Delta| < 1$, $B_1 > 0$, and $B_2 > 0$. The load and source stability circle centers and radii were:

$$c_L = 2.978\angle 51.75^o\,, \quad r_L = 2.131$$

$$c_G = 3.098\angle 162.24^o\,, \quad r_G = 2.254$$

Because both D_1 and D_2 are positive, both stability regions will be the portion of the Smith chart that lies outside the stability circles. For 2 GHz, we find:

$$[K,\mu_1,|\Delta|,B_1,B_2,D_1,D_2] = [1.089,\ 1.056,\ 0.103,\ 1.025,\ 0.954,\ 0.201,\ 0.166]$$

$$c_L = 2.779\angle 50.12^o\,, \quad r_L = 1.723$$

$$c_G = 2.473\angle{-159.36^o}\,, \quad r_G = 1.421$$

The transistor is stable at 2 GHz, with both load and source stability circles being completely outside the unit-circle. □

Problem 13.2 presents an example for which the D_2 parameter is negative, so that the stability regions will be the insides of the stability circles. At one frequency, the unit-circle is partially overlapping with the stability circle, while at another frequency, it lies entirely within the stability circle.

13.6 Power Gains

The amplification (or attenuation) properties of the two-port can be deduced by comparing the power $P_{\rm in}$ going into the two-port to the power P_L coming out of the two-port and going into the load. These were given in Eq. (13.2.1) and we rewrite them as:

$$\begin{aligned} P_{\rm in} &= \frac{1}{2}\,{\rm Re}[V_1^* I_1] = \frac{1}{2}R_{\rm in}|I_1|^2 \qquad \text{(power into two-port)} \\ P_L &= \frac{1}{2}\,{\rm Re}[V_2^* I_2] = \frac{1}{2}R_L|I_2|^2 \qquad \text{(power out of two-port and into load)} \end{aligned} \tag{13.6.1}$$

where we used $V_1 = Z_{\rm in}I_1$, $V_2 = Z_L I_2$, and defined the real parts of the input and load impedances by $R_{\rm in} = {\rm Re}(Z_{\rm in})$ and $R_L = {\rm Re}(Z_L)$. Using the equivalent circuits of Fig. 13.4.1, we may write I_1, I_2 in terms of the generator voltage V_G and obtain:

$$\begin{aligned} P_{\rm in} &= \frac{1}{2}\,\frac{|V_G|^2 R_{\rm in}}{|Z_{\rm in}+Z_G|^2} \\ P_L &= \frac{1}{2}\,\frac{|V_{\rm th}|^2 R_L}{|Z_{\rm out}+Z_L|^2} = \frac{1}{2}\,\frac{|V_G|^2 R_L |Z_{21}|^2}{\big|(Z_{11}+Z_G)(Z_{\rm out}+Z_L)\big|^2} \end{aligned} \tag{13.6.2}$$

Using the identities of Problem 13.1, P_L can also be written in the alternative forms:

$$P_L = \frac{1}{2}\frac{|V_G|^2 R_L |Z_{21}|^2}{\big|(Z_{22}+Z_L)(Z_{\rm in}+Z_G)\big|^2} = \frac{1}{2}\frac{|V_G|^2 R_L|Z_{21}|^2}{\big|(Z_{11}+Z_G)(Z_{22}+Z_L)-Z_{12}Z_{21}\big|^2} \tag{13.6.3}$$

The *maximum* power that can be delivered by the generator to a connected load is called the *available power* of the generator, $P_{\rm avG}$, and is obtained when the load is conjugate-matched to the generator, that is, $P_{\rm avG} = P_{\rm in}$ when $Z_{\rm in} = Z_G^*$.

Similarly, the *available* power from the two-port network, $P_{\rm avN}$, is the maximum power that can be delivered by the Thévenin-equivalent circuit of Fig. 13.4.1 to a connected load, that is, $P_{\rm avN} = P_L$ when $Z_L = Z_{\rm th}^* = Z_{\rm out}^*$. It follows then from Eq. (13.6.2) that the available powers will be:

$$\begin{aligned} P_{\rm avG} &= \max P_{\rm in} = \frac{|V_G|^2}{8R_G} \qquad \text{(available power from generator)} \\ P_{\rm avN} &= \max P_L = \frac{|V_{\rm th}|^2}{8R_{\rm out}} \qquad \text{(available power from network)} \end{aligned} \tag{13.6.4}$$

Using Eq. (13.4.8), $P_{\rm avN}$ can also be written as:

$$P_{\rm avN} = \frac{|V_G|^2}{8R_{\rm out}}\left|\frac{Z_{21}}{Z_{11}+Z_G}\right|^2 \tag{13.6.5}$$

The powers can be expressed completely in terms of the S-parameters of the two-port and the input and output reflection coefficients. With the help of the identities of Problem 13.1, we find the alternative expressions for $P_{\rm in}$ and P_L:

$$\begin{aligned} P_{\rm in} &= \frac{|V_G|^2}{8Z_0}\frac{(1-|\Gamma_{\rm in}|^2)|1-\Gamma_G|^2}{|1-\Gamma_{\rm in}\Gamma_G|^2} \\ P_L &= \frac{|V_G|^2}{8Z_0}\frac{(1-|\Gamma_L|^2)|1-\Gamma_G|^2|S_{21}|^2}{\big|(1-\Gamma_{\rm in}\Gamma_G)(1-S_{22}\Gamma_L)\big|^2} \\ &= \frac{|V_G|^2}{8Z_0}\frac{(1-|\Gamma_L|^2)|1-\Gamma_G|^2|S_{21}|^2}{\big|(1-\Gamma_{\rm out}\Gamma_L)(1-S_{11}\Gamma_G)\big|^2} \\ &= \frac{|V_G|^2}{8Z_0}\frac{(1-|\Gamma_L|^2)|1-\Gamma_G|^2|S_{21}|^2}{\big|(1-S_{11}\Gamma_G)(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_G\Gamma_L\big|^2} \end{aligned} \tag{13.6.6}$$

Similarly, we have for $P_{\rm avG}$ and $P_{\rm avN}$:

$$\begin{aligned} P_{\rm avG} &= \frac{|V_G|^2}{8Z_0}\frac{|1-\Gamma_G|^2}{1-|\Gamma_G|^2} \\ P_{\rm avN} &= \frac{|V_G|^2}{8Z_0}\frac{|1-\Gamma_G|^2|S_{21}|^2}{(1-|\Gamma_{\rm out}|^2)|1-S_{11}\Gamma_G|^2} \end{aligned} \tag{13.6.7}$$

It is evident that $P_{\rm avG}$, $P_{\rm avN}$ are obtained from $P_{\rm in}$, P_L by setting $\Gamma_{\rm in} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\rm out}^*$, which are equivalent to the conjugate-match conditions.

Three widely used definitions for the *power gain* of the two-port network are the *transducer* power gain G_T, the *available* power gain G_a, and the power gain G_p, also called the operating gain. They are defined as follows:

$$\begin{aligned} G_T &= \frac{\text{power out of network}}{\text{maximum power in}} = \frac{P_L}{P_{\rm avG}} \qquad \text{(transducer power gain)} \\ G_a &= \frac{\text{maximum power out}}{\text{maximum power in}} = \frac{P_{\rm avN}}{P_{\rm avG}} \qquad \text{(available power gain)} \\ G_p &= \frac{\text{power out of network}}{\text{power into network}} = \frac{P_L}{P_{\rm in}} \qquad \text{(operating power gain)} \end{aligned} \tag{13.6.8}$$

Each gain is expressible either in terms of the Z-parameters of the two-port, or in terms of its S-parameters. In terms of Z-parameters, the transducer gain is given by the following forms, obtained from the three forms of P_L in Eqs. (13.6.2) and (13.6.3):

$$\begin{aligned} G_T &= \frac{4R_GR_L|Z_{21}|^2}{\big|(Z_{22}+Z_L)(Z_{\rm in}+Z_G)\big|^2} \\ &= \frac{4R_GR_L|Z_{21}|^2}{\big|(Z_{11}+Z_G)(Z_{\rm out}+Z_L)\big|^2} \\ &= \frac{4R_GR_L|Z_{21}|^2}{\big|(Z_{11}+Z_G)(Z_{22}+Z_L)-Z_{12}Z_{21}\big|^2} \end{aligned} \tag{13.6.9}$$

And, in terms of the S-parameters:

$$\boxed{\begin{aligned} G_T &= \frac{1-|\Gamma_G|^2}{|1-\Gamma_{\rm in}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \\ &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-\Gamma_{\rm out}\Gamma_L|^2} \\ &= \frac{(1-|\Gamma_G|^2)|S_{21}|^2(1-|\Gamma_L|^2)}{\big|(1-S_{11}\Gamma_G)(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_G\Gamma_L\big|^2} \end{aligned}} \tag{13.6.10}$$

Similarly, we have for G_a and G_p:

$$\boxed{\begin{aligned} G_a &= \frac{R_G}{R_{\rm out}}\left|\frac{Z_{21}}{Z_{11}+Z_G}\right|^2 = \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1}{1-|\Gamma_{\rm out}|^2} \\ G_p &= \frac{R_L}{R_{\rm in}}\left|\frac{Z_{21}}{Z_{22}+Z_L}\right|^2 = \frac{1}{1-|\Gamma_{\rm in}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \end{aligned}} \tag{13.6.11}$$

The transducer gain G_T is, perhaps, the most representative measure of gain for the two-port because it incorporates the effects of both the load and generator impedances, whereas G_a depends only on the generator impedance and G_p only on the load impedance.

If the generator and load impedances are matched to the reference impedance Z_0, so that $Z_G = Z_L = Z_0$ and $\Gamma_G = \Gamma_L = 0$, and $\Gamma_{\rm in} = S_{11}$, $\Gamma_{\rm out} = S_{22}$, then the power gains reduce to:

$$G_T = |S_{21}|^2\,,\quad G_a = \frac{|S_{21}|^2}{1-|S_{22}|^2}\,,\quad G_p = \frac{|S_{21}|^2}{1-|S_{11}|^2} \tag{13.6.12}$$

A *unilateral* two-port has by definition zero reverse transmission coefficient, that is, $S_{12} = 0$. In this case, the input and output reflection coefficients simplify into:

$$\Gamma_{\text{in}} = S_{11}\,,\quad \Gamma_{\text{out}} = S_{22} \qquad \text{(unilateral two-port)} \tag{13.6.13}$$

The expressions of the power gains simplify somewhat in this case:

$$\begin{aligned} G_{Tu} &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}\,|S_{21}|^2\,\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \\ G_{au} &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}\,|S_{21}|^2\,\frac{1}{1-|S_{22}|^2} \qquad \text{(unilateral gains)} \\ G_{pu} &= \frac{1}{1-|S_{11}|^2}\,|S_{21}|^2\,\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \end{aligned} \tag{13.6.14}$$

For both the bilateral and unilateral cases, the gains G_a, G_p are obtainable from G_T by setting $\Gamma_L = \Gamma_{\text{out}}^*$ and $\Gamma_{\text{in}} = \Gamma_G^*$, respectively, as was the case for P_{avN} and P_{avG}.

The relative power ratios $P_{\text{in}}/P_{\text{avG}}$ and P_L/P_{avN} measure the mismatching between the generator and the two-port and between the load and the two-port. Using the definitions for the power gains, we obtain the input and output *mismatch factors*:

$$M_{\text{in}} = \frac{P_{\text{in}}}{P_{\text{avG}}} = \frac{G_T}{G_p} = \frac{4R_{\text{in}}R_G}{|Z_{\text{in}}+Z_G|^2} = \frac{\big(1-|\Gamma_{\text{in}}|^2\big)\big(1-|\Gamma_G|^2\big)}{|1-\Gamma_{\text{in}}\Gamma_G|^2} \tag{13.6.15}$$

$$M_{\text{out}} = \frac{P_L}{P_{\text{avN}}} = \frac{G_T}{G_a} = \frac{4R_{\text{out}}R_L}{|Z_{\text{out}}+Z_L|^2} = \frac{\big(1-|\Gamma_{\text{out}}|^2\big)\big(1-|\Gamma_L|^2\big)}{|1-\Gamma_{\text{out}}\Gamma_L|^2} \tag{13.6.16}$$

The mismatch factors are always less than or equal to unity (for positive R_{in} and R_{out}.) Clearly, $M_{\text{in}} = 1$ under the conjugate-match condition $Z_{\text{in}} = Z_G^*$ or $\Gamma_{\text{in}} = \Gamma_G^*$, and $M_{\text{out}} = 1$ if $Z_L = Z_{\text{out}}^*$ or $\Gamma_L = \Gamma_{\text{out}}^*$. The mismatch factors can also be written in the following forms, which show more explicitly the mismatch properties:

$$M_{\text{in}} = 1 - \left|\frac{\Gamma_{\text{in}}-\Gamma_G^*}{1-\Gamma_{\text{in}}\Gamma_G}\right|^2\,,\quad M_{\text{out}} = 1 - \left|\frac{\Gamma_{\text{out}}-\Gamma_L^*}{1-\Gamma_{\text{out}}\Gamma_L}\right|^2 \tag{13.6.17}$$

These follow from the identity:

$$|1-\Gamma_1\Gamma_2|^2 - |\Gamma_1-\Gamma_2^*|^2 = \big(1-|\Gamma_1|^2\big)\big(1-|\Gamma_2|^2\big) \tag{13.6.18}$$

The transducer gain is *maximized* when the two-port is *simultaneously* conjugate matched, that is, when $\Gamma_{\text{in}} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\text{out}}^*$. Then, $M_{\text{in}} = M_{\text{out}} = 1$ and the three gains become equal. The common maximum gain achieved by simultaneous matching is called the *maximum available gain* (MAG):

$$G_{T,\max} = G_{a,\max} = G_{p,\max} = G_{\text{MAG}} \tag{13.6.19}$$

Simultaneous matching is discussed in Sec. 13.8. The necessary and sufficient condition for simultaneous matching is $K \geq 1$, where K is the Rollett stability factor. It can be shown that the MAG can be expressed as:

$$\boxed{G_{\text{MAG}} = \frac{|S_{21}|}{|S_{12}|}\big(K - \sqrt{K^2-1}\big)} \qquad \text{(maximum available gain)} \tag{13.6.20}$$

The *maximum stable gain* (MSG) is the maximum value G_{MAG} can have, which is achievable when $K = 1$:

$$\boxed{G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|}} \qquad \text{(maximum stable gain)} \tag{13.6.21}$$

In the unilateral case, the MAG is obtained either by setting $\Gamma_G = \Gamma_{\text{in}}^* = S_{11}^*$ and $\Gamma_L = \Gamma_{\text{out}}^* = S_{22}^*$ in Eq. (13.6.14), or by a careful limiting process in Eq. (13.6.20), in which $K \to \infty$ so that both the numerator factor $K - \sqrt{K^2-1}$ and the denominator factor $|S_{12}|$ tend to zero. With either method, we find the unilateral MAG:

$$G_{\text{MAG},u} = \frac{|S_{21}|^2}{\big(1-|S_{11}|^2\big)\big(1-|S_{22}|^2\big)} = G_1|S_{21}|^2G_2 \qquad \text{(unilateral MAG)} \tag{13.6.22}$$

The maximum unilateral input and output gain factors are:

$$G_1 = \frac{1}{1-|S_{11}|^2}\,,\quad G_2 = \frac{1}{1-|S_{22}|^2} \tag{13.6.23}$$

They are the maxima of the input and output gain factors in Eq. (13.6.14) realized with conjugate matching, that is, with $\Gamma_G = S_{11}^*$ and $\Gamma_L = S_{22}^*$. For any other values of the reflection coefficients (such that $|\Gamma_G| < 1$ and $\Gamma_L| < 1$), we have the following inequalities, which follow from the identity (13.6.18):

$$\frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2} \leq \frac{1}{1-|S_{11}|^2}\,,\qquad \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \leq \frac{1}{1-|S_{22}|^2} \tag{13.6.24}$$

Often two-ports, such as most microwave transistor amplifiers, are approximately unilateral, that is, the measured S-parameters satisfy $|S_{12}| \ll |S_{21}|$. To decide whether the two-port should be treated as unilateral, a figure of merit is used, which is essentially the comparison of the maximum unilateral gain to the transducer gain of the actual device under the same matching conditions, that is, $\Gamma_G = S_{11}^*$ and $\Gamma_L = S_{22}^*$.

For these matched values of Γ_G, Γ_L, the ratio of the bilateral and unilateral transducer gains can be shown to have the form:

$$g_u = \frac{G_T}{G_{Tu}} = \frac{1}{|1-U|^2}\,,\qquad U = \frac{S_{12}S_{21}S_{11}^*S_{22}^*}{\big(1-|S_{11}|^2\big)\big(1-|S_{22}|^2\big)} \tag{13.6.25}$$

The quantity $|U|$ is known as the *unilateral figure of merit.* If the relative gain ratio g_u is near unity (typically, within 10 percent of unity), the two-port may be treated as unilateral.

The MATLAB function `sgain` computes the transducer, available, and operating power gains, given the S-parameters and the reflection coefficients Γ_G, Γ_L. In addition, it computes the unilateral gains, the maximum available gain, and the maximum stable gain. It also computes the unilateral figure of merit ratio (13.6.25). It has usage:

```
Gt = sgain(S,gG,gL);          transducer power gain at given ΓG,ΓL
Ga = sgain(S,gG,'a');         available power gain at given ΓG with ΓL = Γ*out
Gp = sgain(S,gL,'p');         operating power gain at given ΓL with ΓG = Γ*in

Gmag = sgain(S);              maximum available gain (MAG)
Gmsg = sgain(S,'msg');        maximum stable gain (MSG)

Gu = sgain(S,'u');            maximum unilateral gain, Eq. (13.6.22)
G1 = sgain(S,'ui');           maximum unilateral input gain, Eq. (13.6.23)
G2 = sgain(S,'uo');           maximum unilateral output gain, Eq. (13.6.23)
gu = sgain(S,'ufm');          unilateral figure of merit gain ratio, Eq. (13.6.25)
```

The MATLAB functions `gin` and `gout` compute the input and output reflection coefficients from S and Γ_G, Γ_L. They have usage:

```
Gin  = gin(S,gL);             input reflection coefficient, Eq. (13.4.3)
Gout = gout(S,gG);            output reflection coefficient, Eq. (13.4.6)
```

Example 13.6.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor with the following S-parameters at 2 GHz [1324]:

$$S_{11} = 0.61\angle 165^{\circ}\,, \quad S_{21} = 3.72\angle 59^{\circ}\,, \quad S_{12} = 0.05\angle 42^{\circ}\,, \quad S_{22} = 0.45\angle -48^{\circ}$$

Calculate the input and output reflection coefficients and the various power gains, if the amplifier is connected to a generator and load with impedances $Z_G = 10 - 20j$ and $Z_L = 30 + 40j$ ohm.

Solution: The following MATLAB code will calculate all the required gains:

```
Z0 = 50;                              % normalization impedance
ZG = 10+20j;  gG = z2g(ZG,Z0);        % ΓG  =  −0.50 + 0.50j  =  0.71∠135°
ZL = 30-40j;  gL = z2g(ZL,Z0);        % ΓL  =  −0.41 − 0.43j  =  0.59∠−133.15°

S = smat([0.61 165 3.72 59 0.05 42 0.45 -48]);      % reshape S into matrix

Gin = gin(S,gL);                      % Γin  =  0.54∠162.30°
Gout = gout(S,gG);                    % Γout  =  0.45∠−67.46°

Gt = sgain(S,gG,gL);                  % GT  =  4.71, or, 6.73 dB
Ga = sgain(S,gG,'a');                 % Ga  =  11.44, or, 10.58 dB
Gp = sgain(S,gL,'p');                 % Gp  =  10.51, or, 10.22 dB

Gu = sgain(S,'u');                    % Gu  =  27.64, or, 14.41 dB
G1 = sgain(S,'ui');                   % G1  =  1.59, or, 2.02 dB
G2 = sgain(S,'uo');                   % G2  =  1.25, or, 0.98 dB
gu = sgain(S,'ufm');                  % gu  =  1.23, or, 0.89 dB

Gmag = sgain(S);                      % GMAG  =  41.50, or, 16.18 dB
Gmsg = sgain(S,'msg');                % GMSG  =  74.40, or, 18.72 dB
```

The amplifier cannot be considered to be unilateral as the unilateral figure of merit ratio $g_u = 1.23$ is fairly large (larger than 10 percent from unity.)

The amplifier is operating at a gain of $G_T = 6.73$ dB, which is far from the maximum value of $G_{\text{MAG}} = 16.18$ dB. This is because it is mismatched with the given generator and load impedances.

To realize the optimum gain G_{MAG} the amplifier must 'see' certain optimum generator and load impedances or reflection coefficients. These can be calculated by the MATLAB function `smatch` and are found to be:

$$\begin{aligned} \Gamma_G &= 0.82\angle -162.67^{\circ} && \Rightarrow \quad Z_G = \text{g2z}(Z_G, Z_0) = 5.12 - 7.54j\ \Omega \\ \Gamma_L &= 0.75\angle 52.57^{\circ} && \Rightarrow \quad Z_L = \text{g2z}(Z_L, Z_0) = 33.66 + 91.48j\ \Omega \end{aligned}$$

The design of such optimum matching terminations and the function `smatch` are discussed in Sec. 13.8. The functions `g2z` and `z2g` were discussed in Sec. 10.7 . □

13.7 Generalized S-Parameters and Power Waves

The practical usefulness of the S-parameters lies in the fact that the definitions (13.1.4) represent forward and backward traveling waves, which can be measured remotely by connecting a network analyzer to the two-port with transmission lines of characteristic impedance equal to the normalization impedance Z_0. This was depicted in Fig. 13.1.3.

A generalized definition of S-parameters and wave variables can be given by using in Eq. (13.1.4) two different normalization impedances for the input and output ports.

Anticipating that the two-port will be connected to a generator and load of impedances Z_G and Z_L, a particularly convenient choice is to use Z_G for the input normalization impedance and Z_L for the output one, leading to the definition of the *power waves* (as opposed to traveling waves) [958–960,962]:

$$\boxed{\begin{aligned} a_1' &= \frac{V_1 + Z_G I_1}{2\sqrt{R_G}} \\ b_1' &= \frac{V_1 - Z_G^* I_1}{2\sqrt{R_G}} \end{aligned}} \qquad \boxed{\begin{aligned} a_2' &= \frac{V_2 - Z_L I_2}{2\sqrt{R_L}} \\ b_2' &= \frac{V_2 + Z_L^* I_2}{2\sqrt{R_L}} \end{aligned}} \qquad \text{(power waves)} \tag{13.7.1}$$

We note that the b-waves involve the complex-conjugates of the impedances. The quantities R_G, R_L are the resistive parts of Z_G, Z_L and are assumed to be positive. These definitions reduce to the conventional traveling ones if $Z_G = Z_L = Z_0$.

These "wave" variables can no longer be interpreted as incoming and outgoing waves from the two sides of the two-port. However, as we see below, they have a nice interpretation in terms of power transfer to and from the two-port and simplify the expressions for the power gains. Inverting Eqs. (13.7.1), we have:

$$\boxed{\begin{aligned} V_1 &= \frac{1}{\sqrt{R_G}}\,(Z_G^* a_1' + Z_G b_1') \\ I_1 &= \frac{1}{\sqrt{R_G}}\,(a_1' - b_1') \end{aligned}} \qquad \boxed{\begin{aligned} V_2 &= \frac{1}{\sqrt{R_L}}\,(Z_L^* a_2' + Z_L b_2') \\ I_2 &= \frac{1}{\sqrt{R_L}}\,(b_2' - a_2') \end{aligned}} \tag{13.7.2}$$

The power waves can be related directly to the traveling waves. For example, expressing Eqs. (13.7.1) and (13.1.5) in matrix form, we have for port-1:

$$\begin{bmatrix} a_1' \\ b_1' \end{bmatrix} = \frac{1}{2\sqrt{R_G}} \begin{bmatrix} 1 & Z_G \\ 1 & -Z_G^* \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}, \quad \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \frac{1}{\sqrt{Z_0}} \begin{bmatrix} Z_0 & Z_0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$$

It follows that:

$$\begin{bmatrix} a_1' \\ b_1' \end{bmatrix} = \frac{1}{2\sqrt{R_G Z_0}} \begin{bmatrix} 1 & Z_G \\ 1 & -Z_G^* \end{bmatrix} \begin{bmatrix} Z_0 & Z_0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \quad \text{or,}$$

$$\begin{bmatrix} a_1' \\ b_1' \end{bmatrix} = \frac{1}{2\sqrt{R_G Z_0}} \begin{bmatrix} Z_0 + Z_G & Z_0 - Z_G \\ Z_0 - Z_G^* & Z_0 + Z_G^* \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \tag{13.7.3}$$

The entries of this matrix can be expressed directly in terms of the reflection coefficient Γ_G. Using the identities of Problem 13.3, we may rewrite Eq. (13.7.3) and its inverse as follows::

$$\begin{aligned} \begin{bmatrix} a_1' \\ b_1' \end{bmatrix} &= \frac{1}{\sqrt{1 - |\Gamma_G|^2}} \begin{bmatrix} e^{j\phi_G} & -\Gamma_G e^{j\phi_G} \\ -\Gamma_G^* e^{-j\phi_G} & e^{-j\phi_G} \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \\ \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} &= \frac{1}{\sqrt{1 - |\Gamma_G|^2}} \begin{bmatrix} e^{-j\phi_G} & \Gamma_G e^{j\phi_G} \\ \Gamma_G^* e^{-j\phi_G} & e^{j\phi_G} \end{bmatrix} \begin{bmatrix} a_1' \\ b_1' \end{bmatrix} \end{aligned} \tag{13.7.4}$$

where, noting that the quantity $|1 - \Gamma_G|/(1 - \Gamma_G)$ is a pure phase factor, we defined:

$$\Gamma_G = \frac{Z_G - Z_0}{Z_G + Z_0}, \quad e^{j\phi_G} = \frac{|1 - \Gamma_G|}{1 - \Gamma_G} = \frac{1 - \Gamma_G^*}{|1 - \Gamma_G|} \tag{13.7.5}$$

Similarly, we have for the power and traveling waves at port-2:

$$\begin{aligned} \begin{bmatrix} a_2' \\ b_2' \end{bmatrix} &= \frac{1}{\sqrt{1 - |\Gamma_L|^2}} \begin{bmatrix} e^{j\phi_L} & -\Gamma_L e^{j\phi_L} \\ -\Gamma_L^* e^{-j\phi_L} & e^{-j\phi_L} \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \\ \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} &= \frac{1}{\sqrt{1 - |\Gamma_L|^2}} \begin{bmatrix} e^{-j\phi_L} & \Gamma_L e^{j\phi_L} \\ \Gamma_L^* e^{-j\phi_L} & e^{j\phi_L} \end{bmatrix} \begin{bmatrix} a_2' \\ b_2' \end{bmatrix} \end{aligned} \tag{13.7.6}$$

where

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}, \quad e^{j\phi_L} = \frac{|1 - \Gamma_L|}{1 - \Gamma_L} = \frac{1 - \Gamma_L^*}{|1 - \Gamma_L|} \tag{13.7.7}$$

The generalized S-parameters are the scattering parameters with respect to the power wave variables, that is,

$$\begin{bmatrix} b_1' \\ b_2' \end{bmatrix} = \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \begin{bmatrix} a_1' \\ a_2' \end{bmatrix} \quad \Rightarrow \quad \mathbf{b}' = S'\mathbf{a}' \tag{13.7.8}$$

To relate S' to the conventional scattering matrix S, we define the following diagonal matrices:

$$\Gamma = \begin{bmatrix} \Gamma_G & 0 \\ 0 & \Gamma_L \end{bmatrix}, \quad F = \begin{bmatrix} \dfrac{e^{j\phi_G}}{\sqrt{1 - |\Gamma_G|^2}} & 0 \\ 0 & \dfrac{e^{j\phi_L}}{\sqrt{1 - |\Gamma_L|^2}} \end{bmatrix} = \begin{bmatrix} F_G & 0 \\ 0 & F_L \end{bmatrix} \tag{13.7.9}$$

Using these matrices, it follows from Eqs. (13.7.4) and (13.7.6):

$$\begin{aligned} a_1' &= F_G(a_1 - \Gamma_G b_1) \\ a_2' &= F_L(a_2 - \Gamma_L b_2) \end{aligned} \quad \Rightarrow \quad \mathbf{a}' = F(\mathbf{a} - \Gamma\mathbf{b}) \tag{13.7.10}$$

$$\begin{aligned} b_1' &= F_G^*(b_1 - \Gamma_G^* a_1) \\ b_2' &= F_L^*(b_2 - \Gamma_L^* a_2) \end{aligned} \quad \Rightarrow \quad \mathbf{b}' = F^*(\mathbf{b} - \Gamma^*\mathbf{a}) \tag{13.7.11}$$

Using $\mathbf{b} = S\mathbf{a}$, we find

$$\begin{aligned} \mathbf{a}' &= F(\mathbf{a} - \Gamma\mathbf{b}) = F(I - \Gamma S)\mathbf{a} \quad \Rightarrow \quad \mathbf{a} = (I - \Gamma S)^{-1}F^{-1}\mathbf{a}' \\ \mathbf{b}' &= F^*(S - \Gamma^*)\mathbf{a} = F^*(S - \Gamma^*)(I - \Gamma S)^{-1}F^{-1}\mathbf{a}' = S'\mathbf{a}' \end{aligned}$$

where I is the 2×2 unit matrix. Thus, the generalized S-matrix is:

$$\boxed{S' = F^*(S - \Gamma^*)(I - \Gamma S)^{-1}F^{-1}} \tag{13.7.12}$$

We note that $S' = S$ when $Z_G = Z_L = Z_0$, that is, when $\Gamma_G = \Gamma_L = 0$. The explicit expressions for the matrix elements of S' can be derived as follows:

$$\begin{aligned} S_{11}' &= \frac{(S_{11} - \Gamma_G^*)(1 - S_{22}\Gamma_L) + S_{21}S_{12}\Gamma_L}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-2j\phi_G} \\ S_{22}' &= \frac{(S_{22} - \Gamma_L^*)(1 - S_{11}\Gamma_G) + S_{21}S_{12}\Gamma_G}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-2j\phi_L} \end{aligned} \tag{13.7.13a}$$

$$\begin{aligned} S_{21}' &= \frac{\sqrt{1 - |\Gamma_G|^2}\, S_{21}\sqrt{1 - |\Gamma_L|^2}}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-j(\phi_G + \phi_L)} \\ S_{12}' &= \frac{\sqrt{1 - |\Gamma_L|^2}\, S_{12}\sqrt{1 - |\Gamma_G|^2}}{(1 - S_{11}\Gamma_G)(1 - S_{22}\Gamma_L) - S_{12}S_{21}\Gamma_G\Gamma_L}\, e^{-j(\phi_L + \phi_G)} \end{aligned} \tag{13.7.13b}$$

The S_{11}', S_{22}' parameters can be rewritten in terms of the input and output reflection coefficients by using Eq. (13.13.2) and the following factorization identities:

$$\begin{aligned} (S_{11} - \Gamma_G^*)(1 - S_{22}\Gamma_L) + S_{21}S_{12}\Gamma_L &= (\Gamma_{\text{in}} - \Gamma_G^*)(1 - S_{22}\Gamma_L) \\ (S_{22} - \Gamma_L^*)(1 - S_{11}\Gamma_G) + S_{21}S_{12}\Gamma_G &= (\Gamma_{\text{out}} - \Gamma_L^*)(1 - S_{11}\Gamma_G) \end{aligned}$$

It then follows from Eq. (13.7.13) that:

$$S_{11}' = \frac{\Gamma_{\text{in}} - \Gamma_G^*}{1 - \Gamma_{\text{in}}\Gamma_G}\, e^{-2j\phi_G}, \quad S_{22}' = \frac{\Gamma_{\text{out}} - \Gamma_L^*}{1 - \Gamma_{\text{out}}\Gamma_L}\, e^{-2j\phi_L} \tag{13.7.14}$$

Therefore, the mismatch factors (13.6.17) are recognized to be:

$$M_G = 1 - |S'_{11}|^2\,,\quad M_L = 1 - |S'_{22}|^2 \tag{13.7.15}$$

The power flow relations (13.2.1) into and out of the two-port are also valid in terms of the power wave variables. Using Eq. (13.7.2), it can be shown that:

$$\begin{aligned} P_{\text{in}} &= \frac{1}{2}\,\text{Re}[V_1^* I_1] = \frac{1}{2}|a'_1|^2 - \frac{1}{2}|b'_1|^2 \\ P_L &= \frac{1}{2}\,\text{Re}[V_2^* I_2] = \frac{1}{2}|b'_2|^2 - \frac{1}{2}|a'_2|^2 \end{aligned} \tag{13.7.16}$$

In the definitions (13.7.1), the impedances Z_G, Z_L are arbitrary normalization parameters. However, if the two-port is actually connected to a generator V_G with impedance Z_G and a load Z_L, then the power waves take particularly simple forms.

It follows from Fig. 13.1.4 that $V_G = V_1 + Z_G I_1$ and $V_2 = Z_L I_2$. Therefore, definitions Eq. (13.7.1) give:

$$\begin{aligned} a'_1 &= \frac{V_1 + Z_G I_1}{2\sqrt{R_G}} = \frac{V_G}{2\sqrt{R_G}} \\ a'_2 &= \frac{V_2 - Z_L I_2}{2\sqrt{R_L}} = 0 \\ b'_2 &= \frac{V_2 + Z_L^* I_2}{2\sqrt{R_L}} = \frac{Z_L + Z_L^*}{2\sqrt{R_L}}\,I_2 = \frac{2R_L}{2\sqrt{R_L}}\,I_2 = \sqrt{R_L}\,I_2 \end{aligned} \tag{13.7.17}$$

It follows that the available power from the generator and the power delivered to the load are given simply by:

$$\begin{aligned} P_{\text{avG}} &= \frac{|V_G|^2}{8R_G} = \frac{1}{2}|a'_1|^2 \\ P_L &= \frac{1}{2}R_L|I_2|^2 = \frac{1}{2}|b'_2|^2 \end{aligned} \tag{13.7.18}$$

Because $a'_2 = 0$, the generalized scattering matrix gives, $b'_1 = S'_{11}a'_1$ and $b'_2 = S'_{21}a'_1$. The power expressions (13.7.16) then become:

$$\begin{aligned} P_{\text{in}} &= \frac{1}{2}|a'_1|^2 - \frac{1}{2}|b'_1|^2 = \big(1 - |S'_{11}|^2\big)\frac{1}{2}|a'_1|^2 = \big(1 - |S'_{11}|^2\big)P_{\text{avG}} \\ P_L &= \frac{1}{2}|b'_2|^2 - \frac{1}{2}|a'_2|^2 = \frac{1}{2}|b'_2|^2 = |S'_{21}|^2\frac{1}{2}|a'_1|^2 = |S'_{21}|^2 P_{\text{avG}} \end{aligned} \tag{13.7.19}$$

It follows that the transducer and operating power gains are:

$$G_T = \frac{P_L}{P_{\text{avG}}} = |S'_{21}|^2\,,\quad G_p = \frac{P_L}{P_{\text{in}}} = \frac{|S'_{21}|^2}{1 - |S'_{11}|^2} \tag{13.7.20}$$

These also follow from the explicit expressions (13.7.13) and Eqs. (13.6.10) and (13.6.11). We can also express the available power gain in terms of the generalized S-parameters, that is, $G_a = |S'_{21}|^2/(1 - |S'_{22}|^2)$. Thus, we summarize:

$$\boxed{G_T = |S'_{21}|^2\,,\quad G_a = \frac{|S'_{21}|^2}{1 - |S'_{22}|^2}\,,\quad G_p = \frac{|S'_{21}|^2}{1 - |S'_{11}|^2}} \tag{13.7.21}$$

When the load and generator are matched to the network, that is, $\Gamma_{\text{in}} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\text{out}}^*$, the generalized reflections coefficients vanish, $S'_{11} = S'_{22} = 0$, making all the gains equal to each other.

13.8 Simultaneous Conjugate Matching

We saw that the transducer, available, and operating power gains become equal to the maximum available gain G_{MAG} when both the generator and the load are conjugately matched to the two-port, that is, $\Gamma_{\text{in}} = \Gamma_G^*$ and $\Gamma_L = \Gamma_{\text{out}}^*$. Using Eq. (13.5.8), these conditions read explicitly:

$$\begin{aligned} \Gamma_G^* &= S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} = \frac{S_{11} - \Delta\Gamma_L}{1 - S_{22}\Gamma_L} \\ \Gamma_L^* &= S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1 - S_{22}\Gamma_G} = \frac{S_{22} - \Delta\Gamma_G}{1 - S_{11}\Gamma_G} \end{aligned} \tag{13.8.1}$$

Assuming a bilateral two-port, Eqs. (13.8.1) can be solved in the two unknowns Γ_G, Γ_L (eliminating one of the unknowns gives a quadratic equation for the other.) The resulting solutions can be expressed in terms of the parameters (13.5.1):

$$\boxed{\begin{aligned} \Gamma_G &= \frac{B_1 \mp \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \\ \Gamma_L &= \frac{B_2 \mp \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \end{aligned}} \qquad \text{(simultaneous conjugate match)} \tag{13.8.2}$$

where the minus signs are used when $B_1 > 0$ and $B_2 > 0$, and the plus signs, otherwise.

A necessary and sufficient condition for these solutions to have magnitudes $|\Gamma_G| < 1$ and $|\Gamma_L| < 1$ is that the Rollett stability factor be greater than unity, $K > 1$. This is satisfied when the two-port is unconditionally stable, which implies that $K > 1$ and $B_1 > 0, B_2 > 0$.

A conjugate match exists also when the two-port is potentially unstable, but with $K > 1$. Necessarily, this means that $B_1 < 0$, $B_2 < 0$, and also $|\Delta| > 1$. Such cases are rare in practice. For example, most microwave transistors have either $K > 1$ and are stable, or, they are potentially unstable with $K < 1$ and $|\Delta| < 1$.

If the two-port is unilateral, $S_{12} = 0$, then the two equations (13.8.1) decouple, so that the optimum conjugately matched terminations are:

$$\boxed{\Gamma_G = S_{11}^*\,,\quad \Gamma_L = S_{22}^*} \qquad \text{(unilateral conjugate match)} \tag{13.8.3}$$

The MATLAB function `smatch` implements Eqs. (13.8.2). It works only if $K > 1$. Its usage is as follows:

```
[gG,gL] = smatch(S);          % conjugate matched terminations ΓG, ΓL
```

To realize such optimum conjugately matched terminations, matching networks must be used at the input and output of the two-port as shown in Fig. 13.8.1.

The input matching network can be thought as being effectively connected to the impedance $Z_{\text{in}} = Z_G^*$ at its output terminals. It must transform Z_{in} into the actual impedance of the connected generator, typically, $Z_0 = 50$ ohm.

The output matching network must transform the actual load impedance, here Z_0, into the optimum load impedance $Z_L = Z_{\text{out}}^*$.

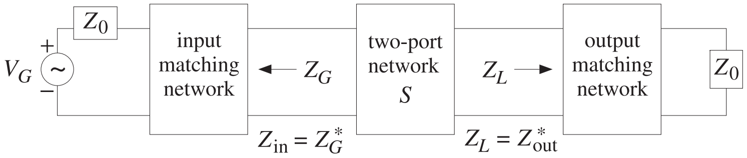


Fig. 13.8.1 Input and output matching networks.

The matching networks may be realized in several possible ways, as discussed in Chap. 12. Stub matching, quarter-wavelength matching, or lumped L-section or Π-section networks may be used. In designing the matching networks, it proves convenient to first design the reverse network as mentioned in Sec. 12.13.

Fig. 13.8.2 shows the procedure for designing the output matching network using a reversed stub matching transformer or a reversed quarter-wave transformer with a parallel stub. In both cases the reversed network is designed to transform the load impedance Z_L^* into Z_0.

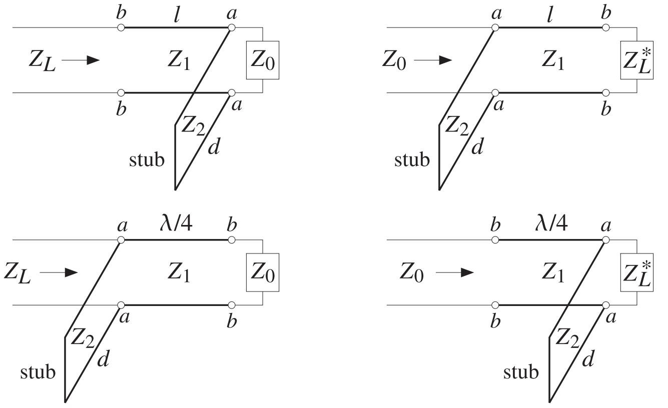


Fig. 13.8.2 Two types of output matching networks and their reversed networks..

Example 13.8.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor having S-parameters at 2 GHz [1324]:

$$S_{11} = 0.61\angle 165^{\circ}\,, \quad S_{21} = 3.72\angle 59^{\circ}\,, \quad S_{12} = 0.05\angle 42^{\circ}\,, \quad S_{22} = 0.45\angle{-48^{\circ}}$$

Determine the optimum conjugately matched source and load terminations, and design appropriate input and output matching networks.

Solution: This is the continuation of Example 13.6.1. The transistor is stable with $K = 1.1752$ and $|\Delta| = 0.1086$. The function `smatch` gives:

$$[\Gamma_G, \Gamma_L] = \text{smatch}(S) \quad \Rightarrow \quad \Gamma_G = 0.8179\angle{-162.6697^{\circ}}\,, \quad \Gamma_L = 0.7495\angle 52.5658^{\circ}$$

The corresponding source, load, input, and output impedances are (with $Z_0 = 50$):

$$Z_G = Z_{\text{in}}^* = 5.1241 - 7.5417j\ \Omega\,, \quad Z_L = Z_{\text{out}}^* = 33.6758 + 91.4816j\ \Omega$$

The locations of the optimum reflection coefficients on the Smith chart are shown in Fig. 13.8.3. For comparison, the unilateral solutions of Eq. (13.8.3) are also shown.

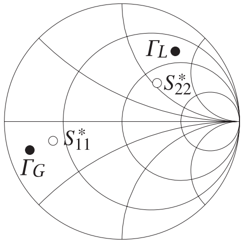


Fig. 13.8.3 Optimum load and source reflection coefficients.

We consider three types of matching networks: (a) microstrip single-stub matching networks with open shunt stubs, shown in Fig. 13.8.4, (b) microstrip quarter-wavelength matching networks with open $\lambda/8$ or $3\lambda/8$ stubs, shown in Fig. 13.8.5, and (c) L-section matching networks, shown in 13.8.6.

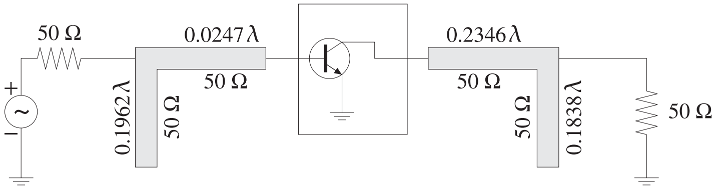


Fig. 13.8.4 Input and output stub matching networks.

In Fig. 13.8.4, the input stub must transform Z_{in} to Z_0. It can be designed with the help of the function `stub1`, which gives the two solutions:

$$dl = \text{stub1}(Z_{\text{in}}/Z_0, \texttt{'po'}) = \begin{bmatrix} 0.3038 & 0.4271 \\ 0.1962 & 0.0247 \end{bmatrix}$$

We choose the lower one, which has the shortest lengths. Thus, the stub length is $d = 0.1962\lambda$ and the segment length $l = 0.0247\lambda$. Both segments can be realized with microstrips of characteristic impedance $Z_0 = 50$ ohm. Similarly, the output matching network can be designed by:

$$dl = \text{stub1}(Z_{\text{out}}/Z_0, \texttt{'po'}) = \begin{bmatrix} 0.3162 & 0.1194 \\ 0.1838 & 0.2346 \end{bmatrix}$$

Again, we choose the lower solutions, $d = 0.1838\lambda$ and $l = 0.2346\lambda$. The solutions using shorted shunt stubs are:

$$\text{stub1}(Z_{\text{in}}/Z_0) = \begin{bmatrix} 0.0538 & 0.4271 \\ 0.4462 & 0.0247 \end{bmatrix}, \quad \text{stub1}(Z_{\text{out}}/Z_0) = \begin{bmatrix} 0.0662 & 0.1194 \\ 0.4338 & 0.2346 \end{bmatrix}$$

Using microstrip lines with alumina substrate ($\epsilon_r = 9.8$), we obtain the following values for the width-to-height ratio, effective permittivity, and wavelength:

$$u = \frac{w}{h} = \text{mstripr}(\epsilon_r, Z_0) = 0.9711$$

$$\epsilon_{\text{eff}} = \text{mstripa}(\epsilon_r, u) = 6.5630$$

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_{\text{eff}}}} = 5.8552 \text{ cm}$$

where $\lambda_0 = 15$ cm is the free-space wavelength at 2 GHz. It follows that the actual segment lengths are $d = 1.1486$ cm, $l = 0.1447$ cm for the input network, and $d = 1.0763$ cm, $l = 1.3734$ cm for the output network.

In the quarter-wavelength method shown in Fig. 13.8.5, we use the function `qwt2` to carry out the design of the required impedances of the microstrip segments. We have for the input and output networks:

$$[Z_1, Z_2] = \text{qwt2}(Z_{\text{in}}, Z_0) = [28.4817, -11.0232]\ \Omega$$

$$[Z_1, Z_2] = \text{qwt2}(Z_{\text{out}}, Z_0) = [118.7832, 103.8782]\ \Omega$$

For the input case, we find $Z_2 = -11.0232\ \Omega$, which means that we should use either a $3\lambda/8$-shorted stub or a $\lambda/8$-opened one. We choose the latter. Similarly, for the output case, we have $Z_2 = 103.8782\ \Omega$, and we choose a $3\lambda/8$-opened stub. The parameters of each microstrip segment are:

$$\begin{array}{lllll}
Z_1 = 28.4817\ \Omega, & u = 2.5832, & \epsilon_{\text{eff}} = 7.2325, & \lambda = 5.578 \text{ cm}, & \lambda/4 = 1.394 \text{ cm} \\
Z_2 = 11.0232\ \Omega, & u = 8.9424, & \epsilon_{\text{eff}} = 8.2974, & \lambda = 5.207 \text{ cm}, & \lambda/8 = 0.651 \text{ cm} \\
Z_1 = 118.7832\ \Omega, & u = 0.0656, & \epsilon_{\text{eff}} = 5.8790, & \lambda = 6.186 \text{ cm}, & \lambda/4 = 1.547 \text{ cm} \\
Z_2 = 103.8782\ \Omega, & u = 0.1169, & \epsilon_{\text{eff}} = 7.9503, & \lambda = 6.149 \text{ cm}, & 3\lambda/8 = 2.306 \text{ cm}
\end{array}$$

Finally, the designs using *L*-sections shown in Fig. 13.8.6, can be carried out with the help of the function `lmatch`. We have the dual solutions for the input and output networks:

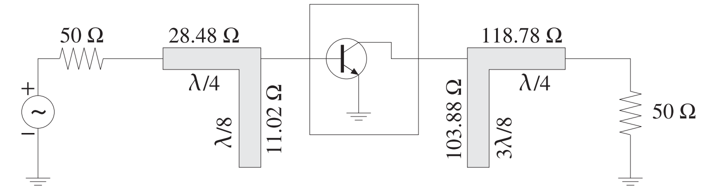


Fig. 13.8.5 Quarter-wavelength matching networks with $\lambda/8$-stubs.

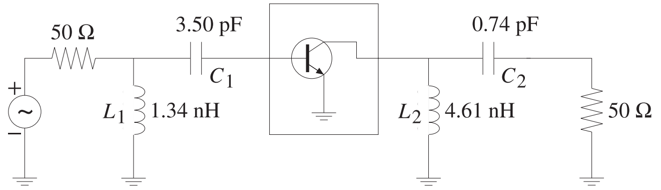


Fig. 13.8.6 Input and output matching with *L*-sections.

$$[X_1, X_2] = \text{lmatch}(Z_0, Z_{\text{in}}, \texttt{'n'}) = \begin{bmatrix} 16.8955 & -22.7058 \\ -16.8955 & 7.6223 \end{bmatrix}$$

$$[X_1, X_2] = \text{lmatch}(Z_{\text{out}}, Z_0, \texttt{'n'}) = \begin{bmatrix} 57.9268 & -107.7472 \\ 502.4796 & 7.6223 \end{bmatrix}$$

According to the usage of `lmatch`, the output network transforms Z_0 into Z_{out}^*, but that is equal to Z_L as required.

Choosing the first rows as the solutions in both cases, the shunt part X_1 will be inductive and the series part X_2, capacitive. At 2 GHz, we find the element values:

$$L_1 = \frac{X_1}{\omega} = 1.3445 \text{ nH}, \quad C_1 = -\frac{1}{\omega X_2} = 3.5047 \text{ pF}$$

$$L_2 = \frac{X_1}{\omega} = 4.6097 \text{ nH}, \quad C_2 = -\frac{1}{\omega X_2} = 0.7386 \text{ pF}$$

The output network, but not the input one, also admits a reversed *L*-section solution:

$$[X_1, X_2] = \text{lmatch}(Z_{\text{out}}, Z_0, \texttt{'r'}) = \begin{bmatrix} 71.8148 & 68.0353 \\ -71.8148 & 114.9280 \end{bmatrix}$$

The essential MATLAB code used to generate the above results was as follows:

```
Z0 = 50; f = 2; w=2*pi*f; la0 = 30/f; er = 9.8;      % f in GHz

S = smat([0.61 165 3.72 59 0.05 42 0.45 -48]);       % S-matrix

[gG,gL] = smatch(S);                            % simultaneous conjugate match
```

```
smith;                                          % draw Fig. 13.8.3
plot(gG, ’.’); plot(conj(S(1,1)), ’o’);
plot(gL, ’.’); plot(conj(S(2,2)), ’o’);

ZG = g2z(gG,Z0);  Zin  = conj(ZG);
ZL = g2z(gL,Z0);  Zout = conj(ZL);

dl = stub1(Zin/Z0, ’po’);                       % single-stub design
dl = stub1(Zout/Z0, ’po’);

u = mstripr(er,Z0);                             % microstrip w/h ratio
eff = mstripa(er,u);                            % effective permittivity
la = la0/sqrt(eff);                             % wavelength within microstrip

[Z1,Z2] = qwt2(Zin, Z0);                        % quarter-wavelength with λ/8 stub
[Z1,Z2] = qwt2(Zout, Z0);

X12 = lmatch(Z0,Zin,’n’);  L1 = X12(1,1)/w;  C1 = -1/(w * X12(1,2))*1e3;
X12 = lmatch(Zout,Z0,’n’); L2 = X12(1,1)/w;  C2 = -1/(w * X12(1,2))*1e3;
X12 = lmatch(Zout,Z0,’r’);                       % L,C in units of nH and pF
```

One could replace the stubs with balanced stubs, as discussed in Sec. 12.9, or use Π- or T-sections instead of L-sections. □

13.9 Power Gain Circles

For a stable two-port, the maximum transducer gain is achieved at single pair of points Γ_G, Γ_L. When the gain G is required to be less than $G_{\rm MAG}$, there will be many possible pairs Γ_G, Γ_L at which the gain G is realized. The locus of such points Γ_G and Γ_L on the Γ-plane is typically a circle of the form:

$$|\Gamma - c| = r \tag{13.9.1}$$

where c, r are the center and radius of the circle and depend on the desired value of the gain G.

In practice, several types of such circles are used, such as unilateral, operating, and available power gain circles, as well as constant noise figure circles, constant SWR circles, and others.

The gain circles allow one to select appropriate values for Γ_G, Γ_L that, in addition to providing the desired gain, also satisfy other requirements, such as striking a balance between minimizing the noise figure and maximizing the gain.

The MATLAB function `sgcirc` calculates the stability circles as well as the operating, available, and unilateral gain circles. Its complete usage is:

```
[c,r] = sgcirc(S,’s’);          % source stability circle
[c,r] = sgcirc(S,’l’);          % load stability circle
[c,r] = sgcirc(S,’p’,G);        % operating power gain circle
[c,r] = sgcirc(S,’a’,G);        % available power gain circle
[c,r] = sgcirc(S,’ui’,G);       % unilateral input gain circle
[c,r] = sgcirc(S,’uo’,G);       % unilateral output gain circle
```

where in the last four cases G is the desired gain in dB.

13.10 Unilateral Gain Circles

We consider only the unconditionally stable unilateral case, which has $|S_{11}| < 1$ and $|S_{22}| < 1$. The dependence of the transducer power gain on Γ_G and Γ_L decouples and the value of the gain may be adjusted by separately choosing Γ_G and Γ_L. We have from Eq. (13.6.14):

$$G_T = \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}\,|S_{21}|^2\,\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G_G\,|S_{21}|^2\,G_L \tag{13.10.1}$$

The input and output gain factors G_G, G_L satisfy the inequalities (13.6.24). Concentrating on the output gain factor, the corresponding gain circle is obtained as the locus of points Γ_L that will lead to a fixed value, say $G_L = G$, which necessarily must be less than the maximum G_2 given in Eq. (13.6.23), that is,

$$\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G \le G_2 = \frac{1}{1-|S_{22}|^2} \tag{13.10.2}$$

Normalizing the gain G to its maximum value $g = G/G_2 = G\big(1-|S_{22}|^2\big)$, we may rewrite (13.10.2) in the form:

$$\frac{\big(1-|\Gamma_L|^2\big)\big(1-|S_{22}|^2\big)}{|1-S_{22}\Gamma_L|^2} = g \le 1 \tag{13.10.3}$$

This equation can easily be rearranged into the equation of a circle $|\Gamma_L - c| = r$, with center and radius given by:

$$\boxed{c = \frac{gS_{22}^*}{1-(1-g)|S_{22}|^2}\,, \quad r = \frac{\sqrt{1-g}\,\big(1-|S_{22}|^2\big)}{1-(1-g)|S_{22}|^2}} \tag{13.10.4}$$

When $g = 1$ or $G = G_2$, the gain circle collapses onto a single point, that is, the optimum point $\Gamma_L = S_{22}^*$. Similarly, we find for the constant gain circles of the input gain factor:

$$\boxed{c = \frac{gS_{11}^*}{1-(1-g)|S_{11}|^2}\,, \quad r = \frac{\sqrt{1-g}\,\big(1-|S_{11}|^2\big)}{1-(1-g)|S_{11}|^2}} \tag{13.10.5}$$

where here, $g = G/G_1 = G\big(1-|S_{11}|^2\big)$ and the circles are $|\Gamma_G - c| = r$.

Both sets of c, r satisfy the conditions $|c| < 1$ and $|c| + r < 1$, the latter implying that the circles lie entirely within the unit circle $|\Gamma| < 1$, that is, within the Smith chart.

Example 13.10.1: A unilateral microwave transistor has S-parameters:

$$S_{11} = 0.8\angle 120^{\rm o}, \quad S_{21} = 4\angle 60^{\rm o}, \quad S_{12} = 0, \quad S_{22} = 0.2\angle{-30}^{\rm o}$$

The unilateral MAG and the maximum input and output gains are obtained as follows:

$$
\begin{aligned}
G_{\text{MAG},u} &= \text{sgain}(S,\texttt{'u'}) = 16.66\ \text{dB}\\
G_1 &= \text{sgain}(S,\texttt{'ui'}) = 4.44\ \text{dB}\\
G_2 &= \text{sgain}(S,\texttt{'uo'}) = 0.18\ \text{dB}
\end{aligned}
$$

Most of the gain is accounted for by the factor $|S_{21}|^2$, which is 12.04 dB. The constant input gain circles for $G_G = 1, 2, 3$ dB are shown in Fig. 13.10.1. Their centers lie along the ray to S_{11}^*. For example, the center and radius of the 3-dB case were computed by

$$[c_3, r_3] = \text{sgcirc}(S,\texttt{'ui'},3) \quad\Rightarrow\quad c_3 = 0.701\angle{-120^\circ}\,,\quad r_3 = 0.233$$

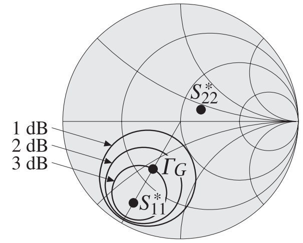


Fig. 13.10.1 Unilateral input gain circles.

Because the output does not provide much gain, we may choose the optimum value $\Gamma_L = S_{22}^* = 0.2\angle 30^\circ$. Then, with any point Γ_G along the 3-dB input gain circle the total transducer gain will be in dB:

$$G_T = G_G + |S_{21}|^2 + G_L = 3 + 12.04 + 0.18 = 15.22\ \text{dB}$$

Points along the 3-dB circle are parametrized as $\Gamma_G = c_3 + r_3 e^{j\phi}$, where ϕ is any angle. Choosing $\phi = \arg(S_{11}^*) - \pi$ will correspond to the point on the circle that lies closest to the origin, that is, $\Gamma_G = 0.468\angle{-120^\circ}$, as shown in Fig. 13.10.1. The corresponding generator and load impedances will be:

$$Z_G = 69.21 + 14.42j\ \Omega, \quad Z_L = 23.15 - 24.02j\ \Omega$$

The MATLAB code used to generate these circles was:

```
S = smat([0.8, 120, 4, 60, 0, 0, 0.2, -30]);

[c1,r1] = sgcirc(S,'ui',1);
[c2,r2] = sgcirc(S,'ui',2);
[c3,r3] = sgcirc(S,'ui',3);

smith;  smithcir(c1,r1);  smithcir(c2,r2);  smithcir(c3,r3);

c = exp(-j*angle(S(1,1))); line([0,real(c)], [0,imag(c)]);

gG = c3 - r3*exp(j*angle(c3));

plot(conj(S(1,1)),'.');  plot(conj(S(2,2)),'.');  plot(gG,'.');
```

The input and output matching networks can be designed using open shunt stubs as in Fig. 13.8.4. The stub lengths are found to be (with $Z_0 = 50\ \Omega$):

$$
\begin{aligned}
dl &= \text{stub1}(Z_G^*/Z_0,\texttt{'po'}) = \begin{bmatrix} 0.3704 & 0.3304\\ 0.1296 & 0.0029 \end{bmatrix}\\
dl &= \text{stub1}(Z_L^*/Z_0,\texttt{'po'}) = \begin{bmatrix} 0.4383 & 0.0994\\ 0.0617 & 0.3173 \end{bmatrix}
\end{aligned}
$$

Choosing the shortest lengths, we have for the input network $d = 0.1296\lambda$, $l = 0.0029\lambda$, and for the output network, $d = 0.0617\lambda$, $l = 0.3173\lambda$. Fig. 13.10.2 depicts the complete matching circuit. □

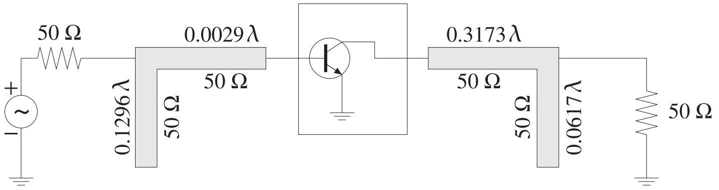


Fig. 13.10.2 Input and output stub matching networks.

13.11 Operating and Available Power Gain Circles

Because the transducer power gain G_T depends on two independent parameters—the source and load reflection coefficients—it is difficult to find the simultaneous locus of points for Γ_G, Γ_L that will result in a given value for the gain.

If the generator is matched, $\Gamma_{\text{in}} = \Gamma_G^*$, then the transducer gain becomes equal to the operating gain $G_T = G_p$ and depends only on the load reflection coefficient Γ_L. The locus of points Γ_L that result in fixed values of G_p are the *operating power gain circles*. Similarly, the *available power gain circles* are obtained by matching the load end, $\Gamma_L = \Gamma_{\text{out}}^*$, and varying Γ_G to achieve fixed values of the available power gain.

Using Eqs. (13.6.11) and (13.5.8), the conditions for achieving a constant value, say G, for the operating or the available power gains are:

$$
\begin{aligned}
G_p &= \frac{1}{1-|\Gamma_{\text{in}}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} = G, \qquad && \Gamma_G^* = \Gamma_{\text{in}} = \frac{S_{11}-\Delta\Gamma_L}{1-S_{22}\Gamma_L}\\
G_a &= \frac{1-|\Gamma_G|^2}{|1-S_{11}\Gamma_G|^2}|S_{21}|^2\frac{1}{1-|\Gamma_{\text{out}}|^2} = G, \qquad && \Gamma_L^* = \Gamma_{\text{out}} = \frac{S_{22}-\Delta\Gamma_G}{1-S_{11}\Gamma_G}
\end{aligned}
\tag{13.11.1}
$$

We consider the operating gain first. Defining the normalized gain $g = G/|S_{21}|^2$, substituting Γ_{in}, and using the definitions (13.5.1), we obtain the condition:

$$
\begin{aligned}
g &= \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2 - |S_{11}-\Delta\Gamma_L|^2} \\
&= \frac{1-|\Gamma_L|^2}{\big(|S_{22}|^2-|\Delta|^2\big)|\Gamma_L|^2 - (S_{22}-\Delta S_{11}^*)\Gamma_L - (S_{22}^*-\Delta^* S_{11})\Gamma_L^* + 1 - |S_{11}|^2} \\
&= \frac{1-|\Gamma_L|^2}{D_2|\Gamma_L|^2 - C_2\Gamma_L - C_2^*\Gamma_L^* + 1 - |S_{11}|^2}
\end{aligned}
$$

This can be rearranged into the form:

$$
|\Gamma_L|^2 - \frac{gC_2}{1+gD_2}\,\Gamma_L - \frac{gC_2^*}{1+gD_2}\,\Gamma_L^* = \frac{1-g\big(1-|S_{11}|^2\big)}{1+gD_2}
$$

and then into the circle form:

$$
\left|\Gamma_L - \frac{gC_2^*}{1+gD_2}\right|^2 = \frac{g^2|C_2|^2}{(1+gD_2)^2} + \frac{1-g\big(1-|S_{11}|^2\big)}{1+gD_2}
$$

Using the identities (13.5.2) and $1-|S_{11}|^2 = 2K|S_{12}S_{21}| + D_2$, which follows from (13.5.1), the right-hand side of the above circle form can be written as:

$$
\frac{g^2|C_2|^2}{(1+gD_2)^2} + \frac{1-g\big(1-|S_{11}|^2\big)}{1+gD_2} = \frac{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}| + 1}{(1+gD_2)^2} \tag{13.11.2}
$$

Thus, the *operating power gain circle* will be $|\Gamma_L - c|^2 = r^2$ with center and radius:

$$
\boxed{c = \frac{gC_2^*}{1+gD_2}\,, \quad r = \frac{\sqrt{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}| + 1}}{|1+gD_2|}} \tag{13.11.3}
$$

The points Γ_L on this circle result into the value $G_p = G$ for the operating gain. Such points can be parametrized as $\Gamma_L = c + re^{j\phi}$, where $0 \le \phi \le 2\pi$. As Γ_L traces this circle, the conjugately matched source coefficient $\Gamma_G = \Gamma_{\text{in}}^*$ will also trace a circle because Γ_{in} is related to Γ_L by the bilinear transformation (13.5.8).

In a similar fashion, we find the *available power gain circles* to be $|\Gamma_G - c|^2 = r^2$, where $g = G/|S_{21}|^2$ and:

$$
\boxed{c = \frac{gC_1^*}{1+gD_1}\,, \quad r = \frac{\sqrt{g^2|S_{12}S_{21}|^2 - 2gK|S_{12}S_{21}| + 1}}{|1+gD_1|}} \tag{13.11.4}
$$

We recall from Sec. 13.5 that the centers of the load and source stability circles were $c_L = C_2^*/D_2$ and $c_G = C_1^*/D_1$. It follows that the centers of the operating power gain circles are along the same ray as c_L, and the centers of the available gain circles are along the same ray as c_G.

For an unconditionally stable two-port, the gain G must be $0 \le G \le G_{\text{MAG}}$, with G_{MAG} given by Eq. (13.6.20). It can be shown easily that the quantities under the square roots in the definitions of the radii r in Eqs. (13.11.3) and (13.11.4) are non-negative. The gain circles lie inside the unit circle for all such values of G. The radii r vanish when $G = G_{\text{MAG}}$, that is, the circles collapse into single points corresponding to the simultaneous conjugate matched solutions of Eq. (13.8.2).

The MATLAB function `sgcirc` calculates the center and radii c, r of the operating and available power gain circles. It has usage, where G must be entered in dB:

```
[c,r] = sgcirc(S,'p',G);          operating power gain circle
[c,r] = sgcirc(S,'a',G);          available power gain circle
```

Example 13.11.1: A microwave transistor amplifier uses the Hewlett-Packard AT-41410 NPN bipolar transistor with the following S-parameters at 2 GHz [1324]:

$$
S_{11} = 0.61\angle 165^{\circ}\,, \quad S_{21} = 3.72\angle 59^{\circ}\,, \quad S_{12} = 0.05\angle 42^{\circ}\,, \quad S_{22} = 0.45\angle{-48^{\circ}}
$$

Calculate G_{MAG} and plot the operating and available power gain circles for $G = 13, 14, 15$ dB. Then, design source and load matching circuits for the case $G = 15$ dB by choosing the reflection coefficient that has the smallest magnitude.

Solution: The MAG was calculated in Example 13.6.1, $G_{\text{MAG}} = 16.18$ dB. The gain circles and the corresponding load and source stability circles are shown in Fig. 13.11.1. The operating gain and load stability circles were computed and plotted by the MATLAB statements:

```
[c1,r1] = sgcirc(S,'p',13);            % c1 = 0.4443∠52.56°, r1 = 0.5212
[c2,r2] = sgcirc(S,'p',14);            % c2 = 0.5297∠52.56°, r2 = 0.4205
[c3,r3] = sgcirc(S,'p',15);            % c3 = 0.6253∠52.56°, r3 = 0.2968
[cL,rL] = sgcirc(S,'l');               % cL = 2.0600∠52.56°, rL = 0.9753

smith; smithcir(cL,rL,1.7);            % display portion of circle with |ΓL| ≤ 1.7
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);
```

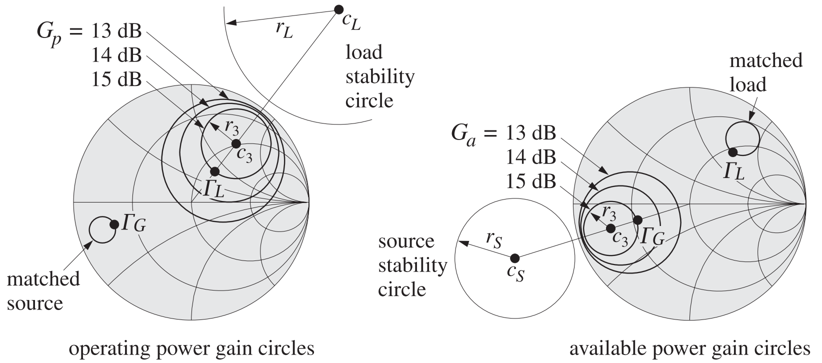


Fig. 13.11.1 Operating and available power gain circles.

The gain circles lie entirely within the unit circle, for example, we have $r_3 + |c_3| = 0.9221 < 1$, and their centers lie along the ray of c_L. As Γ_L traces the 15-dB circle, the corresponding $\Gamma_G = \Gamma_{\text{in}}^*$ traces its own circle, also lying within the unit circle. The following MATLAB code computes and adds that circle to the above Smith chart plots:

```
phi = linspace(0,2*pi,361);           % equally spaced angles at 1° intervals
gammaL = c3 + r3 * exp(j*phi);        % points on 15-dB operating gain circle
gammaG = conj(gin(S,gammaL));         % circle of conjugate matched source points
plot(gammaG);
```

In particular, the point $\varGamma_L$ on the 15-dB circle that lies closest to the origin is $\varGamma_L = c_3 - r_3 e^{j\arg c_3} = 0.3285\angle 52.56^{\rm o}$. The corresponding matched load will be $\varGamma_G = \varGamma_{\rm in}^* = 0.6805\angle{-163.88^{\rm o}}$. These and the corresponding source and load impedances were computed by the MATLAB statements:

```
gL = c3 - r3*exp(j*angle(c3));      zL = g2z(gL);
gG = conj(gin(S,gL));               zG = g2z(gG);
```

The source and load impedances normalized to $Z_0 = 50$ ohm are:

$$z_G = \frac{Z_G}{Z_0} = 0.1938 - 0.1363j\,,\quad z_L = \frac{Z_L}{Z_0} = 1.2590 + 0.7361j$$

The matching circuits can be designed in a variety of ways as in Example 13.8.1. Using open shunt stubs, we can determine the stub and line segment lengths with the help of the function `stub1`:

$$dl = \mathrm{stub1}(z_G^*, \texttt{'po'}) = \begin{bmatrix} 0.3286 & 0.4122 \\ 0.1714 & 0.0431 \end{bmatrix}$$

$$dl = \mathrm{stub1}(z_L^*, \texttt{'po'}) = \begin{bmatrix} 0.4033 & 0.0786 \\ 0.0967 & 0.2754 \end{bmatrix}$$

In both cases, we may choose the lower solutions as they have shorter total length $d + l$. The available power gain circles can be determined in a similar fashion with the help of the MATLAB statements:

```
[c1,r1] = sgcirc(S,'a',13);           % c1 = 0.5384∠−162.67°, r1 = 0.4373
[c2,r2] = sgcirc(S,'a',14);           % c2 = 0.6227∠−162.67°, r2 = 0.3422
[c3,r3] = sgcirc(S,'a',15);           % c3 = 0.7111∠−162.67°, r3 = 0.2337
[cG,rG] = sgcirc(S,'s');              % cG = 1.5748∠−162.67°, rG = 0.5162

smith; smithcir(cG,rG);               % plot entire source stability circle
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);
```

Again, the circles lie entirely within the unit circle. As $\varGamma_G$ traces the 15-dB circle, the corresponding matched load $\varGamma_L = \varGamma_{\rm out}^*$ traces its own circle on the $\varGamma$-plane. It can be plotted with:

```
phi = linspace(0,2*pi,361);           % equally spaced angles at 1° intervals
gammaG = c3 + r3 * exp(j*phi);        % points on 15-dB available gain circle
gammaL = conj(gout(S,gammaG));        % circle of conjugate matched loads
plot(gammaL);
```

In particular, the point $\varGamma_G = c_3 - r_3 e^{j\arg c_3} = 0.4774\angle{-162.67^{\rm o}}$ lies closest to the origin. The corresponding matched load will have $\varGamma_L = \varGamma_{\rm out}^* = 0.5728\angle 50.76^{\rm o}$. The resulting normalized impedances are:

$$z_G = \frac{Z_G}{Z_0} = 0.3609 - 0.1329j\,,\quad z_L = \frac{Z_L}{Z_0} = 1.1135 + 1.4704j$$

and the corresponding stub matching networks will have lengths:

$$\mathrm{stub1}(z_G^*, \texttt{'po'}) = \begin{bmatrix} 0.3684 & 0.3905 \\ 0.1316 & 0.0613 \end{bmatrix},\quad \mathrm{stub1}(z_L^*, \texttt{'po'}) = \begin{bmatrix} 0.3488 & 0.1030 \\ 0.1512 & 0.2560 \end{bmatrix}$$

The lower solutions have the shortest lengths. For both the operating and available gain cases, the stub matching circuits will be similar to those in Fig. 13.8.4. □

When the two-port is potentially unstable (but with $|S_{11}| < 1$ and $|S_{22}| < 1$,) the stability circles intersect with the unit-circle, as shown in Fig. 13.5.2. In this case, the operating and available power gain circles also intersect the unit-circle and at the *same* points as the stability circles.

We demonstrate this in the specific case of $K < 1$, $|S_{11}| < 1$, $|S_{22}| < 1$, but with $D_2 > 0$, an example of which is shown in Fig. 13.11.2. The intersection of an operating gain circle with the unit-circle is obtained by setting $|\varGamma_L| = 1$ in the circle equation $|\varGamma_L - c| = r$. Writing $\varGamma_L = e^{j\theta_L}$ and $c = |c|e^{j\theta_c}$, we have:

$$r^2 = |\varGamma_L - c|^2 = 1 - 2|c|\cos(\theta_L - \theta_c) + |c|^2 \quad\Rightarrow\quad \cos(\theta_L - \theta_c) = \frac{1 + |c|^2 - r^2}{2|c|}$$

Similarly, the intersection of the load stability circle with the unit-circle leads to the relationship:

$$r_L^2 = |\varGamma_L - c_L|^2 = 1 - 2|c_L|\cos(\theta_L - \theta_{c_L}) + |c_L|^2 \quad\Rightarrow\quad \cos(\theta_L - \theta_{c_L}) = \frac{1 + |c_L|^2 - r_L^2}{2|c_L|}$$

Because $c = gC_2^*/(1 + gD_2)$, $c_L = C_2^*/D_2$, and $D_2 > 0$, it follows that the phase angles of c and c_L will be equal, $\theta_c = \theta_{c_L}$. Therefore, in order for the load stability circle and the gain circle to intersect the unit-circle at the same $\varGamma_L = e^{j\theta_L}$, the following condition must be satisfied:

$$\cos(\theta_L - \theta_c) = \frac{1 + |c|^2 - r^2}{2|c|} = \frac{1 + |c_L|^2 - r_L^2}{2|c_L|} \tag{13.11.5}$$

Using the identities $1 - |S_{11}|^2 = B_2 - D_2$ and $1 - |S_{11}|^2 = \big(|c_L|^2 - r_L^2\big)D_2$, which follow from Eqs. (13.5.1) and (13.5.6), we obtain:

$$\frac{1 + |c_L|^2 - r_L^2}{2|c_L|} = \frac{1 + (B_2 - D_2)/D_2}{2|C_2|/|D_2|} = \frac{B_2}{2|C_2|}$$

where we used $D_2 > 0$. Similarly, Eq. (13.11.2) can be written in the form:

$$r^2 = |c|^2 + \frac{1 - g\big(1 - |S_{11}|^2\big)}{1 + gD_2} \quad\Rightarrow\quad |c|^2 - r^2 = \frac{g\big(1 - |S_{11}|^2\big) - 1}{1 + gD_2} = \frac{g(B_2 - D_2) - 1}{1 + gD_2}$$

Therefore, we have:

$$\frac{1+|c|^2-r^2}{2|c|}=\frac{1+\big(g(B_2-D_2)-1\big)/(1+gD_2)}{2g|C_2|/|1+gD_2|}=\frac{B_2}{2|C_2|}$$

Thus, Eq. (13.11.5) is satisfied. This condition has two solutions for θ_L that correspond to the two points of intersection with the unit-circle. When $D_2 > 0$, we have $\arg c = \arg C_2^* = -\arg C_2$. Therefore, the two solutions for $\Gamma_L = e^{j\theta_L}$ will be:

$$\Gamma_L = e^{j\theta_L}\,,\quad \theta_L = -\arg(C_2)\pm\operatorname{acos}\left(\frac{B_2}{2|C_2|}\right) \tag{13.11.6}$$

Similarly, the points of intersection of the unit-circle and the available gain circles and source stability circle are:

$$\Gamma_G = e^{j\theta_G}\,,\quad \theta_G = -\arg(C_1)\pm\operatorname{acos}\left(\frac{B_1}{2|C_1|}\right) \tag{13.11.7}$$

Actually, these expressions work also when $D_2 < 0$ or $D_1 < 0$.

Example 13.11.2: The microwave transistor Hewlett-Packard AT-41410 NPN is potentially unstable at 1 GHz with the following S-parameters [1324]:

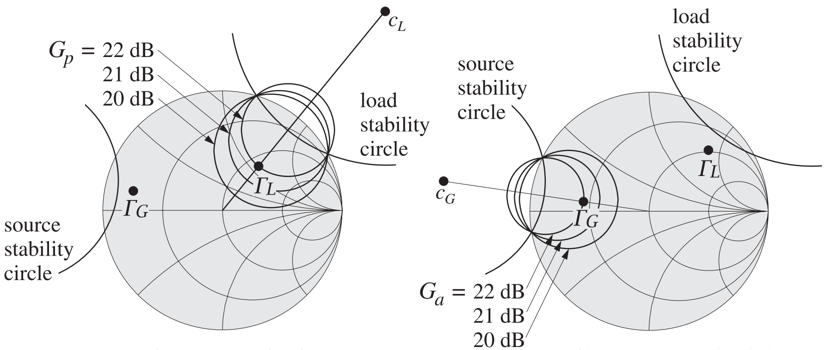
$$S_{11} = 0.6\angle{-163^{\text{o}}}\,,\quad S_{21} = 7.12\angle 86^{\text{o}}\,,\quad S_{12} = 0.039\angle 35^{\text{o}}\,,\quad S_{22} = 0.50\angle{-38^{\text{o}}}$$

Calculate G_{MSG} and plot the operating and available power gain circles for $G = 20, 21, 22$ dB. Then, design source and load matching circuits for the 22-dB case by choosing the reflection coefficients that have the smallest magnitudes.

Solution: The MSG computed from Eq. (13.6.21) is $G_{\text{MSG}} = 22.61$ dB. Fig. 13.11.2 depicts the operating and available power gain circles as well as the load and source stability circles. The stability parameters are: $K = 0.7667, \mu_1 = 0.8643, |\Delta| = 0.1893, D_1 = 0.3242, D_2 = 0.2142$. The computations and plots are done with the following MATLAB code:[†]

```
S = smat([0.60, -163, 7.12, 86, 0.039, 35, 0.50, -38]);      % S-parameters

[K,mu,D,B1,B2,C1,C2,D1,D2] = sparam(S);        % stability parameters

Gmsg = db(sgain(S,'msg'));                   % G_MSG = 22.61 dB
                                             % operating power gain circles:
[c1,r1] = sgcirc(S,'p',20);                  % c_1 = 0.6418∠50.80°, r_1 = 0.4768
[c2,r2] = sgcirc(S,'p',21);                  % c_2 = 0.7502∠50.80°, r_2 = 0.4221
[c3,r3] = sgcirc(S,'p',22);                  % c_3 = 0.8666∠50.80°, r_3 = 0.3893
                                             % load and source stability circles:
[cL,rL] = sgcirc(S,'l');                     % c_L = 2.1608∠50.80°, r_L = 1.2965
[cG,rG] = sgcirc(S,'s');                     % c_G = 1.7456∠171.69°, r_G = 0.8566

smith; smithcir(cL,rL,1.5); smithcir(cG,rG,1.5);        % plot Smith charts
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);      % plot gain circles

gL = c3 - r3*exp(j*angle(c3));               % Γ_L of smallest magnitude
gG = conj(gin(S,gL));                        % corresponding matched Γ_G
plot(gL,'.'); plot(gG,'.');
```

[†] The function `db` converts absolute scales to dB. The function `ab` converts from dB to absolute units.



operating power gain circles

available power gain circles

Fig. 13.11.2 Operating and available power gain circles.

```
                                            % available power gain circles:
[c1,r1] = sgcirc(S,'a',20);                 % c_1 = 0.6809∠171.69°, r_1 = 0.4137
[c2,r2] = sgcirc(S,'a',21);                 % c_2 = 0.7786∠171.69°, r_2 = 0.3582
[c3,r3] = sgcirc(S,'a',22);                 % c_3 = 0.8787∠171.69°, r_3 = 0.3228

figure;
smith; smithcir(cL,rL,1.5); smithcir(cG,rG,1.5);
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3);

gG = c3 - r3*exp(j*angle(c3));              % Γ_G of smallest magnitude
gL = conj(gout(S,gG));                      % corresponding matched Γ_L
plot(gL,'.'); plot(gG,'.');
```

Because $D_1 > 0$ and $D_2 > 0$, the stability regions are the portions of the unit-circle that lie outside the source and load stability circles. We note that the operating gain circles intersect the unit-circle at exactly the same points as the load stability circle, and the available gain circles intersect it at the same points as the source stability circle.

The value of Γ_L on the 22-dB operating gain circle that lies closest to the origin is $\Gamma_L = c_3 - r_3 e^{j\arg c_3} = 0.4773\angle 50.80^{\text{o}}$ and the corresponding matched source is $\Gamma_G = \Gamma_{\text{in}}^* = 0.7632\angle 167.69^{\text{o}}$. We note that both Γ_L and Γ_G lie in their respective stability regions.

For the 22-dB available gain circle (also denoted by c_3, r_3), the closest Γ_G to the origin will be $\Gamma_G = c_3 - r_3 e^{j\arg c_3} = 0.5559\angle 171.69^{\text{o}}$ with a corresponding matched load $\Gamma_L = \Gamma_{\text{out}}^* = 0.7147\angle 45.81^{\text{o}}$. Again, both Γ_L, Γ_G lie in their stable regions.

Once the Γ_G, Γ_L have been determined, the corresponding matching input and output networks can be designed with the methods of Example 13.8.1. □

13.12 Noise Figure Circles

Every device is a source of internally generated noise. The noise entering the device and the internal noise must be added to obtain the total input system noise. If the device is an amplifier, the total system noise power will be amplified at the output by the gain of the device. If the output load is matched, this gain will be the available gain.

The internally generated noise is quantified in practice either by the *effective noise temperature* T_e, or by the *noise figure* F of the device. The internal noise power is given by $P_n = kT_eB$, where k is the Boltzmann constant and B the bandwidth in Hz. These concepts are discussed further in Sec. 15.8. The relationship between T_e and F is defined in terms of a standard reference temperature $T_0 = 290$ K (degrees Kelvin):

$$F = 1 + \frac{T_e}{T_0} \tag{13.12.1}$$

The noise figure is usually quoted in dB, $F_{\text{dB}} = 10\log_{10}F$. Because the available gain of a two-port depends on the source impedance Z_G, or the source reflection coefficient Γ_G, so will the noise figure.

The optimum source impedance $Z_{G\text{opt}}$ corresponds to the minimum noise figure F_{min} that can be achieved by the two-port. For other values of Z_G, the noise figure F is greater than F_{min} and is given by [94–97]:

$$F = F_{\text{min}} + \frac{R_n}{R_G|Z_{G\text{opt}}|^2}\,|Z_G - Z_{G\text{opt}}|^2 \tag{13.12.2}$$

where $R_G = \text{Re}(Z_G)$ and R_n is an equivalent noise resistance. We note that $F = F_{\text{min}}$ when $Z_G = Z_{G\text{opt}}$. Defining the normalized noise resistance $r_n = R_n/Z_0$, where $Z_0 = 50$ ohm, we may write Eq. (13.12.2) in terms of the corresponding source reflection coefficients:

$$\boxed{F = F_{\text{min}} + 4r_n\,\frac{|\Gamma_G - \Gamma_{G\text{opt}}|^2}{|1+\Gamma_{G\text{opt}}|^2\big(1-|\Gamma_G|^2\big)}} \tag{13.12.3}$$

The parameters F_{min}, r_n, and $\Gamma_{G\text{opt}}$ characterize the noise properties of the two-port and are usually known.

In designing low-noise microwave amplifiers, one would want to achieve the minimum noise figure and the maximum gain. Unfortunately, the optimum source reflection coefficient $\Gamma_{G\text{opt}}$ does not necessarily correspond to the maximum available gain.

The noise figure circles and the available gain circles are useful tools that allow one to obtain a balance between low noise and high gain designs. The noise figure circles are the locus of points Γ_G that correspond to fixed values of F. They are obtained by rewriting Eq. (13.12.3) as the equation of a circle $|\Gamma_G - c|^2 = r^2$. We write Eq. (13.12.3) in the form:

$$\frac{|\Gamma_G - \Gamma_{G\text{opt}}|^2}{1-|\Gamma_G|^2} = N\,,\quad \text{where}\quad N = \frac{(F-F_{\text{min}})\,|1+\Gamma_{G\text{opt}}|^2}{4r_n} \tag{13.12.4}$$

which can be rearranged into the circle equation:

$$\left|\Gamma_G - \frac{\Gamma_{G\text{opt}}}{N+1}\right|^2 = \frac{N^2 + N\big(1-|\Gamma_{G\text{opt}}|^2\big)}{(N+1)^2}$$

Thus, the center and radius of the noise figure circle are:

$$\boxed{c = \frac{\Gamma_{G\text{opt}}}{N+1}\,,\quad r = \frac{\sqrt{N^2 + N\big(1-|\Gamma_{G\text{opt}}|^2\big)}}{N+1}} \tag{13.12.5}$$

The MATLAB function `nfcirc` implements Eq. (13.12.5). Its inputs are the noise parameters F_{min}, r_n, $\Gamma_{G\text{opt}}$, and the desired value of F in dB, and its outputs are c, r:

```
[c,r] = nfcirc(F,Fmin,rn,gGopt);        % noise figure circles
```

The function `nfig` implements Eq. (13.12.3). Its inputs are F_{min}, r_n, $\Gamma_{G\text{opt}}$, and a vector of values of Γ_G, and its output is the corresponding vector of values of F:

```
F = nfig(Fmin, rn, gGopt, gG);          % calculate noise figure F in dB
```

Example 13.12.1: The microwave transistor of Example 13.11.1 has the following noise parameters at 2 GHz [1324]: $F_{\text{min}} = 1.6$ dB, $r_n = 0.16$, and $\Gamma_{G\text{opt}} = 0.26\angle 172^{\text{o}}$.

Determine the matched load $\Gamma_{L\text{opt}}$ corresponding to $\Gamma_{G\text{opt}}$ and calculate the available gain. Then, plot the noise figure circles for $F = 1.7, 1.8, 1.9, 2.0$ dB.

For the 1.8-dB noise figure circle, determine Γ_G, Γ_L that correspond to the maximum possible available gain and design appropriate input and output matching networks.

Solution: The conjugate matched load corresponding to $\Gamma_{G\text{opt}}$ is:

$$\Gamma_{L\text{opt}} = \Gamma_{\text{out}}^* = \left[\frac{S_{22} - \Delta\Gamma_{G\text{opt}}}{1 - S_{11}\Gamma_{G\text{opt}}}\right]^* = 0.4927\angle 52.50^{\text{o}}$$

The value of the available gain at $\Gamma_{G\text{opt}}$ is $G_{a,\text{opt}} = 13.66$ dB. This is to be compared with the MAG of 16.18 dB determined in Example 13.11.1. To increase the available gain, we must also increase the noise figure. Fig. 13.12.1 shows the locations of the optimum reflection coefficients, as well as several noise figure circles.

The MATLAB code for generating this graph was:[†]

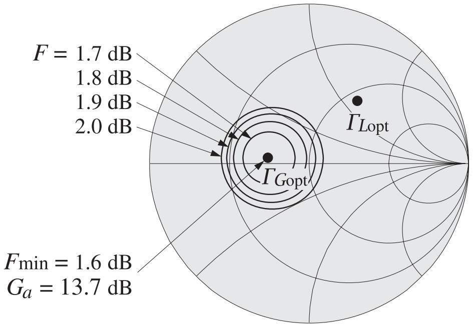


Fig. 13.12.1 Noise figure circles.

```
S = smat([0.61, 165, 3.72, 59, 0.05, 42, 0.45, -48]);

Fmin = 1.6; rn = 0.16; gGopt = p2c(0.26, 172);
```

[†] The function `p2c` converts from phasor form to cartesian complex form, and the function `c2p`, from cartesian to phasor form.

```
Gmag = db(sgain(S,'mag'));                     % maximum available gain
Gaopt = db(sgain(S,gGopt,'a'))                 % available gain at ΓGopt

gLopt = conj(gout(S,gGopt));                   % matched load

[c1,r1] = nfcirc(1.7,Fmin,rn,gGopt);           % noise figure circles
[c2,r2] = nfcirc(1.8,Fmin,rn,gGopt);
[c3,r3] = nfcirc(1.9,Fmin,rn,gGopt);
[c4,r4] = nfcirc(2.0,Fmin,rn,gGopt);

smith; plot([gGopt, gLopt],'.');
smithcir(c1,r1); smithcir(c2,r2); smithcir(c3,r3); smithcir(c4,r4);
```

The larger the noise figure F, the larger the radius of its circle. As F increases, so does the available gain. But as the gain increases, the radius of its circle decreases. Thus, for a fixed value of F, there will be a maximum value of the available gain corresponding to that gain circle that has the smallest radius and is tangent to the noise figure circle.

In the extreme case of the maximum available gain, the available gain circle collapses to a point—the simultaneous conjugate matched point $\Gamma_G = 0.8179\angle{-162.67^{\circ}}$— with a corresponding noise figure of $F = 4.28$ dB. These results can be calculated by the MATLAB statements:

```
gG = smatch(S);
F = nfig(Fmin, rn, gopt, gG);
```

Thus, we see that increasing the gain comes at the price of increasing the noise figure. As Γ_G traces the $F = 1.8$ dB circle, the available gain G_a varies as shown in Fig. 13.12.2. Points around this circle can be parametrized as $\Gamma_G = c_2 + r_2 e^{j\phi}$, with $0 \le \phi \le 2\pi$. Fig. 13.12.2 plots G_a versus the angle ϕ. We note that the gain varies between the limits $12.22 \le G_a \le 14.81$ dB.

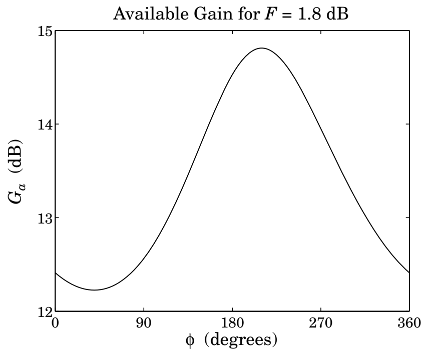


Fig. 13.12.2 Variation of available gain around the noise figure circle $F = 1.8$ dB.

The maximum value, $G_a = 14.81$ dB, is reached when $\Gamma_G = 0.4478\angle{-169.73^{\circ}}$, with a resulting matched load $\Gamma_L = \Gamma_{\text{out}}^* = 0.5574\angle{52.50^{\circ}}$. The two points Γ_G, Γ_L, as well as the $G_a = 14.81$ dB gain circle, which is tangential to the 1.8-dB noise figure circle, are shown in Fig. 13.12.3.

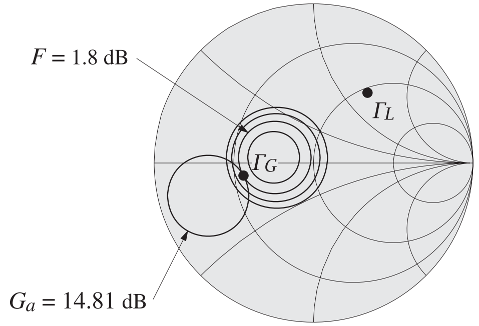


Fig. 13.12.3 Maximum available gain for given noise figure.

The following MATLAB code performs these calculations and plots:

```
phi = linspace(0,2*pi,721);                    % angle in 1/2° increments
gG = c2 + r2*exp(j*phi);                       % ΓG around the c2, r2 circle
G = db(sgain(S,gG,'a'));                       % available gain in dB
plot(phi*180/pi, G);

[Ga,i] = max(G);                               % maximum available gain

gammaG = gG(i);                                % ΓG for maximum gain
gammaL = conj(gout(S,gammaG));                 % matched load ΓL

[ca,ra] = sgcirc(S,'a',Ga);                    % available gain circle

smith; smithcir(c2,r2); smithcir(ca,ra);
plot([gammaG,gammaL],'.');
```

The maximum gain and the point of tangency with the noise figure circle are determined by direct search, that is, evaluating the gain around the 1.8-dB noise figure circle and finding where it reaches a maximum.

The input and output stub matching networks can be designed with the help of the function `stub1`. The normalized source and load impedances are:

$$z_G = \frac{1+\Gamma_G}{1-\Gamma_G} = 0.3840 - 0.0767j\,, \quad z_L = \frac{1+\Gamma_L}{1-\Gamma_L} = 1.0904 + 1.3993j$$

The stub matching networks have lengths:

$$\text{stub1}(z_G^*, \texttt{'po'}) = \begin{bmatrix} 0.3749 & 0.3977 \\ 0.1251 & 0.0738 \end{bmatrix}, \quad \text{stub1}(z_L^*, \texttt{'po'}) = \begin{bmatrix} 0.3519 & 0.0991 \\ 0.1481 & 0.2250 \end{bmatrix}$$

The lower solutions have shorter total lengths $d + l$. The implementation of the matching networks with microstrip lines will be similar to that in Fig. 13.8.4. □

If the two-port is potentially unstable, one must be check that the resulting solutions for Γ_G, Γ_L both lie in their respective stability regions. Problems 13.6 and 13.7 illustrate the design of such potentially unstable low noise microwave amplifiers.

13.13 Problems

13.1 Using the relationships (13.4.3) and (13.4.6), derive the following identities:

$$\begin{aligned}(Z_{11}+Z_G)(Z_{22}+Z_L)-Z_{12}Z_{21} &= \\ (Z_{22}+Z_L)(Z_{\text{in}}+Z_G) &= (Z_{11}+Z_G)(Z_{\text{out}}+Z_L)\end{aligned} \tag{13.13.1}$$

$$\begin{aligned}(1-S_{11}\Gamma_G)(1-S_{22}\Gamma_L)-S_{12}S_{21}\Gamma_G\Gamma_L &= \\ (1-S_{22}\Gamma_L)(1-\Gamma_{\text{in}}\Gamma_G) &= (1-S_{11}\Gamma_G)(1-\Gamma_{\text{out}}\Gamma_L)\end{aligned} \tag{13.13.2}$$

Using Eqs. (13.4.4) and (13.4.5), show that:

$$\frac{Z_{21}}{Z_{22}+Z_L} = \frac{S_{21}}{1-S_{22}\Gamma_L}\,\frac{1-\Gamma_L}{1-\Gamma_{\text{in}}}\,, \qquad \frac{Z_{21}}{Z_{11}+Z_G} = \frac{S_{21}}{1-S_{11}\Gamma_G}\,\frac{1-\Gamma_G}{1-\Gamma_{\text{out}}} \tag{13.13.3}$$

$$\frac{2Z_0}{Z_{\text{in}}+Z_G} = \frac{(1-\Gamma_{\text{in}})(1-\Gamma_G)}{1-\Gamma_{\text{in}}\Gamma_G}\,, \qquad \frac{2Z_0}{Z_{\text{out}}+Z_L} = \frac{(1-\Gamma_{\text{out}})(1-\Gamma_L)}{1-\Gamma_{\text{out}}\Gamma_L} \tag{13.13.4}$$

Finally, for the real part $R_L = \text{Re}(Z_L)$, show that:

$$Z_L = Z_0\frac{1+\Gamma_L}{1-\Gamma_L} \quad \Rightarrow \quad R_L = Z_0\frac{1-|\Gamma_L|^2}{|1-\Gamma_L|^2} \tag{13.13.5}$$

13.2 *Computer Experiment.* The Hewlett-Packard ATF-10136 GaAs FET transistor has the following S-parameters at 4 GHz and 8 GHz [1324]:

$$\begin{array}{llll} S_{11} = 0.54\angle{-120^{\text{o}}}\,, & S_{21} = 3.60\angle 61^{\text{o}}\,, & S_{12} = 0.137\angle 31^{\text{o}}\,, & S_{22} = 0.22\angle{-49^{\text{o}}} \\ S_{11} = 0.60\angle 87^{\text{o}}\,, & S_{21} = 2.09\angle{-32^{\text{o}}}\,, & S_{12} = 0.21\angle{-36^{\text{o}}}\,, & S_{22} = 0.32\angle{-48^{\text{o}}} \end{array}$$

Determine the stability parameters, stability circles, and stability regions at the two frequencies.

13.3 Derive the following relationships, where $R_G = \text{Re}(Z_G)$:

$$\frac{Z_0+Z_G}{2\sqrt{R_G Z_0}} = \frac{1}{\sqrt{1-|\Gamma_G|^2}}\,\frac{|1-\Gamma_G|}{1-\Gamma_G}\,, \qquad \frac{Z_0-Z_G}{2\sqrt{R_G Z_0}} = -\frac{\Gamma_G}{\sqrt{1-|\Gamma_G|^2}}\,\frac{|1-\Gamma_G|}{1-\Gamma_G}$$

13.4 Derive Eqs. (13.7.13) relating the generalized S-parameters of power waves to the conventional S-parameters.

13.5 Derive the expression Eq. (13.6.20) for the maximum available gain G_{MAG}, and show that it is the maximum of all three gains, that is, transducer, available, and operating gains.

13.6 *Computer Experiment.* The microwave transistor of Example 13.11.2 has the following noise parameters at a frequency of 1 GHz [1324]: $F_{\min} = 1.3$ dB, $r_n = 0.16$, and $\Gamma_{G\text{opt}} = 0.06\angle 49^{\text{o}}$.

Determine the matched load $\Gamma_{L\text{opt}}$ corresponding to $\Gamma_{G\text{opt}}$ and calculate the available gain. Then, plot the noise figure circles for $F = 1.4, 1.5, 1.6$ dB.

For the 1.5-dB noise figure circle, determine the values of Γ_G, Γ_L that correspond to the maximum possible available gain.

Design microstrip stub matching circuits for the computed values of Γ_G, Γ_L.

13.7 *Computer Experiment.* The Hewlett-Packard ATF-36163 pseudomorphic high electron mobility transistor (PHEMT) has the following S- and noise parameters at 6 GHz [1324]:

$$S_{11} = 0.75\angle{-131^{\text{o}}}\,, \quad S_{21} = 3.95\angle 55^{\text{o}}\,, \quad S_{12} = 0.13\angle{-12^{\text{o}}}\,, \quad S_{22} = 0.27\angle{-116^{\text{o}}}$$

$$F_{\min} = 0.66 \text{ dB}, \quad r_n = 0.15, \quad \Gamma_{G\text{opt}} = 0.55\angle 88^{\text{o}}$$

Plot the $F = 0.7, 0.8, 0.9$ dB noise figure circles. On the 0.7-dB circle, determine the source reflection coefficient Γ_G that corresponds to maximum available gain, and then determine the corresponding matched load coefficient Γ_L.

Design microstrip stub matching circuits for the computed values of Γ_G, Γ_L.

13.8 *Computer Experiment.* In this experiment, you will carry out two low-noise microwave amplifier designs, including the corresponding input and output matching networks. The first design fixes the noise figure and finds the maximum gain that can be used. The second design fixes the desired gain and finds the minimum noise figure that may be achieved.

The Hewlett-Packard Agilent ATF-34143 PHEMT transistor is suitable for low-noise amplifiers in cellular/PCS base stations, low-earth-orbit and multipoint microwave distribution systems, and other low-noise applications.

At 2 GHz, its S-parameters and noise-figure data are as follows, for biasing conditions of $V_{DS} = 4$ V and $I_{DS} = 40$ mA:

$$\begin{array}{ll} S_{11} = 0.700\angle{-150^{\text{o}}}\,, & S_{12} = 0.081\angle 19^{\text{o}} \\ S_{21} = 6.002\angle 73^{\text{o}}\,, & S_{22} = 0.210\angle{-150^{\text{o}}} \end{array}$$

$$F_{\min} = 0.22 \text{ dB}, \quad r_n = 0.09, \quad \Gamma_{G\text{opt}} = 0.66\angle 67^{\text{o}}$$

a. At 2 GHz, the transistor is potentially unstable. Calculate the stability parameters K, μ, Δ, D_1, D_2. Calculate the MSG in dB.

Draw a basic Smith chart and place on it the source and load stability circles (display only a small portion of each circle outside the Smith chart.)

Then, determine the parts of the Smith chart that correspond to the source and load stability regions.

b. For the given optimum reflection coefficient $\Gamma_{G\text{opt}}$, calculate the corresponding load reflection coefficient $\Gamma_{L\text{opt}}$ assuming a matched load.

Place the two points $\Gamma_{G\text{opt}}, \Gamma_{L\text{opt}}$ on the above Smith chart and determine whether they lie in their respective stability regions.

c. Calculate the available gain $G_{a,\text{opt}}$ in dB that corresponds to $\Gamma_{G\text{opt}}$.

Add the corresponding available gain circle to the above Smith chart. (Note that the source stability circle and the available gain circles intersect the Smith chart at the same points.)

d. Add to your Smith chart the noise figure circles corresponding to the noise figure values of $F = 0.25, 0.30, 0.35$ dB.

For the case $F = 0.35$ dB, calculate and plot the available gain G_a in dB as Γ_G traces the noise-figure circle. Determine the maximum value of G_a and the corresponding value of Γ_G.

Place on your Smith chart the available gain circle corresponding to this maximum G_a. Place also the corresponding point Γ_G, which should be the point of tangency between the gain and noise figure circles.

Calculate and place on the Smith chart the corresponding load reflection coefficient $\Gamma_L = \Gamma_{\text{out}}^*$. Verify that the two points Γ_G, Γ_L lie in their respective stability regions.

In addition, for comparison purposes, place on your Smith chart the available gain circles corresponding to the values $G_a = 15$ and 16 dB.

e. The points Γ_G and Γ_L determined in the previous question achieve the maximum gain for the given noise figure of $F = 0.35$ dB.

Design input and output stub matching networks that match the amplifier to a 50-ohm generator and a 50-ohm load. Use "parallel/open" microstrip stubs having 50-ohm characteristic impedance and alumina substrate of relative permittivity of $\epsilon_r = 9.8$.

Determine the stub lengths d, l in units of λ, the wavelength inside the microstrip lines. Choose always the solution with the shortest total length $d + l$.

Determine the effective permittivity ϵ_{eff} of the stubs, the stub wavelength λ in cm, and the width/height ratio, w/h. Then, determine the stub lengths d, l in cm.

Finally, make a schematic of your final design that shows both the input and output matching networks (as in Fig.10.8.3.)

f. The above design sets $F = 0.35$ dB and finds the maximum achievable gain. Carry out an alternative design as follows. Start with a desired available gain of $G_a = 16$ dB and draw the corresponding available gain circle on your Smith chart.

As Γ_G traces the portion of this circle that lies inside the Smith chart, compute the corresponding noise figure F. (Points on the circle can be parametrized by $\Gamma_G = c + re^{j\phi}$, but you must keep only those that have $|\Gamma_G| < 1$.)

Find the minimum among these values of F in dB and calculate the corresponding value of Γ_G. Calculate the corresponding matched Γ_L.

Add to your Smith chart the corresponding noise figure circle and place on it the points Γ_G and Γ_L.

g. Design the appropriate stub matching networks as in part 13.8.

14

Radiation Fields

14.1 Currents and Charges as Sources of Fields

Here we discuss how a given distribution of currents and charges can generate and radiate electromagnetic waves. Typically, the current distribution is localized in some region of space (for example, currents on a wire antenna.) The current source generates electromagnetic fields, which can propagate to far distances from the source location.

It proves convenient to work with the electric and magnetic *potentials* rather than the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields themselves. Basically, two of Maxwell's equations allow us to introduce these potentials; then, the other two, written in terms of these potentials, take a simple wave-equation form. The two Maxwell equations,

$$\nabla\cdot\boldsymbol{B}=0\,,\quad \nabla\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t} \tag{14.1.1}$$

imply the existence of the magnetic and electric potentials $\boldsymbol{A}(\mathbf{r},t)$ and $\varphi(\mathbf{r},t)$, such that the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ are obtainable by

$$\boxed{\begin{aligned}&\boldsymbol{E}=-\nabla\varphi-\frac{\partial\boldsymbol{A}}{\partial t}\\&\boldsymbol{B}=\nabla\times\boldsymbol{A}\end{aligned}} \tag{14.1.2}$$

Indeed, the divergenceless of $\boldsymbol{B}$ implies the existence of $\boldsymbol{A}$, such that $\boldsymbol{B}=\nabla\times\boldsymbol{A}$. Then, Faraday's law can be written as

$$\nabla\times\boldsymbol{E}=-\frac{\partial\boldsymbol{B}}{\partial t}=-\nabla\times\frac{\partial\boldsymbol{A}}{\partial t}\quad\Rightarrow\quad\nabla\times\big(\boldsymbol{E}+\frac{\partial\boldsymbol{A}}{\partial t}\big)=0$$

Thus, the quantity $\boldsymbol{E}+\partial\boldsymbol{A}/\partial t$ is curl-less and can be represented as the gradient of a scalar potential, that is, $\boldsymbol{E}+\partial\boldsymbol{A}/\partial t=-\nabla\varphi$.

The potentials $\boldsymbol{A}$ and φ are not uniquely defined. For example, they may be changed by adding constants to them. Even more freedom is possible, known as *gauge invariance* of Maxwell's equations. Indeed, for any scalar function $f(\mathbf{r},t)$, the following *gauge transformation* leaves $\boldsymbol{E}$ and $\boldsymbol{B}$ invariant:

$$\boxed{\begin{aligned}\varphi' &= \varphi - \frac{\partial f}{\partial t}\\ \boldsymbol{A}' &= \boldsymbol{A} + \nabla f\end{aligned}} \qquad \text{(gauge transformation)} \tag{14.1.3}$$

For example, we have for the electric field:

$$\boldsymbol{E}' = -\nabla\varphi' - \frac{\partial \boldsymbol{A}'}{\partial t} = -\nabla\big(\varphi - \frac{\partial f}{\partial t}\big) - \frac{\partial}{\partial t}\big(\boldsymbol{A} + \nabla f\big) = -\nabla\varphi - \frac{\partial \boldsymbol{A}}{\partial t} = \boldsymbol{E}$$

This freedom in selecting the potentials allows us to impose some convenient *constraints* between them. In discussing radiation problems, it is customary to impose the *Lorenz condition*:[†]

$$\boxed{\nabla\cdot\boldsymbol{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t} = 0} \qquad \text{(Lorenz condition)} \tag{14.1.4}$$

We will also refer to it as *Lorenz gauge or radiation gauge.* Under the gauge transformation (14.1.3), we have:

$$\nabla\cdot\boldsymbol{A}' + \frac{1}{c^2}\frac{\partial\varphi'}{\partial t} = \big(\nabla\cdot\boldsymbol{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t}\big) - \big(\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \nabla^2 f\big)$$

Therefore, if $\boldsymbol{A}, \varphi$ did not satisfy the constraint (14.1.4), the transformed potentials $\boldsymbol{A}', \varphi'$ could be made to satisfy it by an appropriate choice of the function f, that is, by choosing f to be the solution of the inhomogeneous wave equation:

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \nabla^2 f = \nabla\cdot\boldsymbol{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t}$$

Using Eqs. (14.1.2) and (14.1.4) into the remaining two of Maxwell's equations,

$$\nabla\cdot\boldsymbol{E} = \frac{1}{\epsilon}\rho, \quad \nabla\times\boldsymbol{B} = \mu\boldsymbol{J} + \frac{1}{c^2}\frac{\partial\boldsymbol{E}}{\partial t} \tag{14.1.5}$$

we find,

$$\begin{aligned}\nabla\cdot\boldsymbol{E} = \nabla\cdot\big(-\nabla\varphi - \frac{\partial\boldsymbol{A}}{\partial t}\big) &= -\nabla^2\varphi - \frac{\partial}{\partial t}(\nabla\cdot\boldsymbol{A}) = -\nabla^2\varphi - \frac{\partial}{\partial t}\big(-\frac{1}{c^2}\frac{\partial\varphi}{\partial t}\big)\\ &= \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi\end{aligned}$$

and, similarly,

$$\begin{aligned}\nabla\times\boldsymbol{B} - \frac{1}{c^2}\frac{\partial\boldsymbol{E}}{\partial t} &= \nabla\times(\nabla\times\boldsymbol{A}) - \frac{1}{c^2}\frac{\partial}{\partial t}\big(-\nabla\varphi - \frac{\partial\boldsymbol{A}}{\partial t}\big)\\ &= \nabla\times(\nabla\times\boldsymbol{A}) + \nabla\big(\frac{1}{c^2}\frac{\partial\varphi}{\partial t}\big) + \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2}\end{aligned}$$

$$\begin{aligned}&= \nabla\times(\nabla\times\boldsymbol{A}) - \nabla(\nabla\cdot\boldsymbol{A}) + \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2}\\ &= \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2} - \nabla^2\boldsymbol{A}\end{aligned}$$

where we used the identity $\nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A}$. Therefore, Maxwell's equations (14.1.5) take the equivalent wave-equation forms for the potentials:

$$\boxed{\begin{aligned}\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \nabla^2\varphi &= \frac{1}{\epsilon}\rho\\ \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2} - \nabla^2\boldsymbol{A} &= \mu\boldsymbol{J}\end{aligned}} \qquad \text{(wave equations)} \tag{14.1.7}$$

To summarize, the densities $\rho, \boldsymbol{J}$ may be thought of as the *sources* that generate the potentials $\varphi, \boldsymbol{A}$, from which the fields $\boldsymbol{E}, \boldsymbol{B}$ may be computed via Eqs. (14.1.2).

The Lorenz condition is compatible with Eqs. (14.1.7) and implies charge conservation. Indeed, we have from (14.1.7)

$$\big(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\big)\big(\nabla\cdot\boldsymbol{A} + \frac{1}{c^2}\frac{\partial\varphi}{\partial t}\big) = \mu\nabla\cdot\boldsymbol{J} + \frac{1}{\epsilon c^2}\frac{\partial\rho}{\partial t} = \mu\big(\nabla\cdot\boldsymbol{J} + \frac{\partial\rho}{\partial t}\big)$$

where we used $\mu\epsilon = 1/c^2$. Thus, the Lorenz condition (14.1.4) implies the charge conservation law:

$$\nabla\cdot\boldsymbol{J} + \frac{\partial\rho}{\partial t} = 0 \tag{14.1.8}$$

14.2 Retarded Potentials

The main result that we would like to show here is that if the source densities $\rho, \boldsymbol{J}$ are known, the causal solutions of the wave equations (14.1.7) are given by:

$$\boxed{\begin{aligned}\varphi(\mathbf{r}, t) &= \int_V \frac{\rho(\mathbf{r}', t - \frac{R}{c})}{4\pi\epsilon R}\, d^3\mathbf{r}'\\ \boldsymbol{A}(\mathbf{r}, t) &= \int_V \frac{\mu\boldsymbol{J}(\mathbf{r}', t - \frac{R}{c})}{4\pi R}\, d^3\mathbf{r}'\end{aligned}} \qquad \text{(retarded potentials)} \tag{14.2.1}$$

where $R = |\mathbf{r} - \mathbf{r}'|$ is the distance from the field (observation) point $\mathbf{r}$ to the source point $\mathbf{r}'$, as shown in Fig. 14.2.1. The integrations are over the localized volume V in which the source densities $\rho, \boldsymbol{J}$ are non-zero.

In words, the potential $\varphi(\mathbf{r}, t)$ at a field point $\mathbf{r}$ at time t is obtainable by superimposing the fields due to the infinitesimal charge $\rho(\mathbf{r}', t')\, d^3\mathbf{r}'$ that resided within the volume element $d^3\mathbf{r}'$ at time instant t', which is R/c seconds earlier than t, that is, $t' = t - R/c$.

Thus, in accordance with our intuitive notions of causality, a change at the source point $\mathbf{r}'$ is not felt instantaneously at the field point $\mathbf{r}$, but takes R/c seconds to get there, that is, it propagates with the speed of light. Equations (14.2.1) are referred to

---

[†] Almost universally wrongly attributed to H. A. Lorentz instead of L. V. Lorenz. See Refs. [64–70] for the historical roots of scalar and vector potentials and gauge transformations.

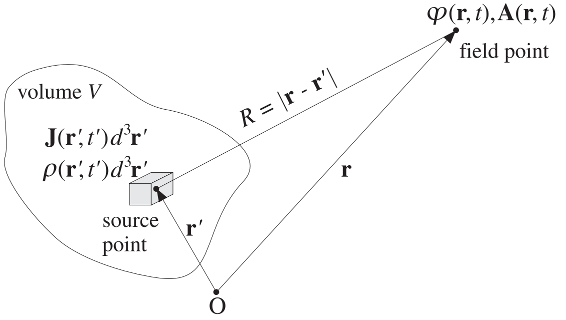

$\varphi(\mathbf{r},t),\mathbf{A}(\mathbf{r},t)$
field point
volume V
$R = |\mathbf{r} - \mathbf{r}'|$
$\mathbf{J}(\mathbf{r}',t')d^3\mathbf{r}'$
$\rho(\mathbf{r}',t')d^3\mathbf{r}'$
$\mathbf{r}$
source point
$\mathbf{r}'$
O

Fig. 14.2.1 Retarded potentials generated by a localized current/charge distribution.

as the *retarded* potentials because the sources inside the integrals are evaluated at the retarded time $t' = t - R/c$.

To prove (14.2.1), we consider first the solution to the following scalar wave equation driven by a time-dependent point source located at the origin:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f(t)\,\delta^{(3)}(\mathbf{r}) \tag{14.2.2}$$

where $f(t)$ is an arbitrary function of time and $\delta^{(3)}(\mathbf{r})$ is the 3-dimensional delta function. We show below that the causal solution of Eq. (14.2.2) is:[†]

$$u(\mathbf{r},t) = \frac{f(t')}{4\pi r} = \frac{f\big(t - \frac{r}{c}\big)}{4\pi r} = f\big(t - \frac{r}{c}\big)g(\mathbf{r}), \qquad \text{where} \quad g(\mathbf{r}) = \frac{1}{4\pi r} \tag{14.2.3}$$

with $t' = t - r/c$ and $r = |\mathbf{r}|$. The function $g(\mathbf{r})$ is recognized as the Green's function for the electrostatic Coulomb problem and satisfies:

$$\boldsymbol{\nabla} g = -\hat{\mathbf{r}}\,\frac{1}{4\pi r^2} = -\hat{\mathbf{r}}\,\frac{g}{r}, \qquad \nabla^2 g = -\delta^{(3)}(\mathbf{r}) \tag{14.2.4}$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$ is the radial unit vector. We note also that because $f(t - r/c)$ depends on r only through its t-dependence, we have:

$$\frac{\partial}{\partial r}f(t - r/c) = -\frac{1}{c}\frac{\partial}{\partial t}f(t - r/c) = -\frac{1}{c}\dot{f}$$

It follows that $\boldsymbol{\nabla} f = -\hat{\mathbf{r}}\dot{f}/c$ and

$$\boldsymbol{\nabla}^2 f = -(\boldsymbol{\nabla}\cdot\hat{\mathbf{r}})\frac{\dot{f}}{c} - \frac{1}{c}\hat{\mathbf{r}}\cdot\boldsymbol{\nabla}\dot{f} = -(\boldsymbol{\nabla}\cdot\hat{\mathbf{r}})\frac{\dot{f}}{c} - \frac{1}{c}\hat{\mathbf{r}}\cdot\big(-\hat{\mathbf{r}}\frac{\ddot{f}}{c}\big) = -\frac{2\dot{f}}{cr} + \frac{1}{c^2}\ddot{f} \tag{14.2.5}$$

where we used the result $\boldsymbol{\nabla}\cdot\hat{\mathbf{r}} = 2/r$.[‡] Using Eqs. (14.2.3)–(14.2.5) into the identity:

$$\nabla^2 u = \nabla^2(fg) = 2\boldsymbol{\nabla} f\cdot\boldsymbol{\nabla} g + g\nabla^2 f + f\nabla^2 g$$

[†] The anticausal, or time-advanced, solution is $u(\mathbf{r},t) = f(t + r/c)g(\mathbf{r})$.

[‡] Indeed, $\boldsymbol{\nabla}\cdot\hat{\mathbf{r}} = \boldsymbol{\nabla}\cdot(\mathbf{r}/r) = (\boldsymbol{\nabla}\cdot\mathbf{r})/r + \mathbf{r}\cdot(-\hat{\mathbf{r}}/r^2) = 3/r - 1/r = 2/r$.

we obtain,

$$\nabla^2 u = 2\big(-\hat{\mathbf{r}}\frac{\dot{f}}{c}\big)\cdot\big(-\hat{\mathbf{r}}\frac{g}{r}\big) - \frac{2\dot{f}}{cr}g + \frac{1}{c^2}\ddot{f}g - f\big(t - \frac{r}{c}\big)\delta^{(3)}(\mathbf{r})$$

The first two terms cancel and the fourth term can be written as $f(t)\delta^{(3)}(\mathbf{r})$ because the delta function forces $\mathbf{r} = 0$. Recognizing that the third term is

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{1}{c^2}\ddot{f}g$$

we have,

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - f(t)\,\delta^{(3)}(\mathbf{r})$$

which shows Eq. (14.2.2). Next, we shift the point source to location $\mathbf{r}'$, and find the solution to the wave equation:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f(\mathbf{r}',t)\,\delta^{(3)}(\mathbf{r} - \mathbf{r}') \qquad \Rightarrow \qquad u(\mathbf{r},t) = \frac{f(\mathbf{r}', t - R/c)}{4\pi R} \tag{14.2.6}$$

where $R = |\mathbf{r} - \mathbf{r}'|$ and we have allowed the function f to also depend on $\mathbf{r}'$. Note that here $\mathbf{r}'$ is fixed and the field point $\mathbf{r}$ is variable.

Using linearity, we may form now the linear combination of several such point sources located at various values of $\mathbf{r}'$ and get the corresponding linear combination of solutions. For example, the sum of two sources will result in the sum of solutions:

$$f(\mathbf{r}_1',t)\,\delta^{(3)}(\mathbf{r} - \mathbf{r}_1') + f(\mathbf{r}_2',t)\,\delta^{(3)}(\mathbf{r} - \mathbf{r}_2') \quad \Rightarrow \quad \frac{f(\mathbf{r}_1', t - R_1/c)}{4\pi R_1} + \frac{f(\mathbf{r}_2', t - R_2/c)}{4\pi R_2}$$

where $R_1 = |\mathbf{r} - \mathbf{r}_1'|$, $R_2 = |\mathbf{r} - \mathbf{r}_2'|$. More generally, integrating over the whole volume V over which $f(\mathbf{r}',t)$ is nonzero, we have for the sum of sources:

$$f(\mathbf{r},t) = \int_V f(\mathbf{r}',t)\,\delta^{(3)}(\mathbf{r} - \mathbf{r}')\,d^3\mathbf{r}'$$

and the corresponding sum of solutions:

$$u(\mathbf{r},t) = \int_V \frac{f(\mathbf{r}', t - R/c)}{4\pi R}\,d^3\mathbf{r}' \tag{14.2.7}$$

where $R = |\mathbf{r} - \mathbf{r}'|$. Thus, this is the causal solution to the general wave equation:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f(\mathbf{r},t) \tag{14.2.8}$$

The retarded potentials (14.2.1) are special cases of Eq. (14.2.7), applied for $f(\mathbf{r},t) = \rho(\mathbf{r},t)/\epsilon$ and $f(\mathbf{r},t) = \mu\mathbf{J}(\mathbf{r},t)$.

14.3 Harmonic Time Dependence

Since we are primarily interested in single-frequency waves, we will Fourier transform all previous results. This is equivalent to assuming a sinusoidal time dependence $e^{j\omega t}$ for all quantities. For example,

$$\varphi(\mathbf{r},t)=\varphi(\mathbf{r})e^{j\omega t}\,,\quad \rho(\mathbf{r},t)=\rho(\mathbf{r})e^{j\omega t}\,,\quad \text{etc.}$$

Then, the retarded solutions (14.2.1) become:

$$\varphi(\mathbf{r})e^{j\omega t}=\int_V \frac{\rho(\mathbf{r}')e^{j\omega(t-\frac{R}{c})}}{4\pi\epsilon R}\,d^3\mathbf{r}'$$

Canceling a common factor $e^{j\omega t}$ from both sides, we obtain for the phasor part of the retarded potentials, where $R=|\mathbf{r}-\mathbf{r}'|$:

$$\boxed{\begin{aligned}\varphi(\mathbf{r})&=\int_V \frac{\rho(\mathbf{r}')e^{-jkR}}{4\pi\epsilon R}\,d^3\mathbf{r}'\\ \boldsymbol{A}(\mathbf{r})&=\int_V \frac{\mu\boldsymbol{J}(\mathbf{r}')e^{-jkR}}{4\pi R}\,d^3\mathbf{r}'\end{aligned}}\,,\qquad \text{where}\quad \boxed{k=\frac{\omega}{c}} \tag{14.3.1}$$

The quantity k represents the free-space wavenumber and is related to the wavelength via $k=2\pi/\lambda$. An alternative way to obtain Eqs. (14.3.1) is to start with the wave equations and replace the time derivatives by $\partial_t\to j\omega$. Equations (14.1.7) become then the Helmholtz equations:

$$\boxed{\begin{aligned}\nabla^2\varphi+k^2\varphi&=-\frac{1}{\epsilon}\rho\\ \nabla^2\boldsymbol{A}+k^2\boldsymbol{A}&=-\mu\boldsymbol{J}\end{aligned}} \tag{14.3.2}$$

Their solutions may be written in the convolutional form:[†]

$$\boxed{\begin{aligned}\varphi(\mathbf{r})&=\int_V \frac{1}{\epsilon}\rho(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,d^3\mathbf{r}'\\ \boldsymbol{A}(\mathbf{r})&=\int_V \mu\boldsymbol{J}(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,d^3\mathbf{r}'\end{aligned}} \tag{14.3.3}$$

where $G(\mathbf{r})$ is the Green's function for the Helmholtz equation:

$$\nabla^2G+k^2G=-\delta^{(3)}(\mathbf{r})\,,\qquad G(\mathbf{r})=\frac{e^{-jkr}}{4\pi r} \tag{14.3.4}$$

Replacing $\partial/\partial t$ by $j\omega$, the Lorenz condition (14.1.4) takes the form:

$$\nabla\cdot\boldsymbol{A}+j\omega\mu\epsilon\varphi=0 \tag{14.3.5}$$

[†] The integrals in (14.3.1) or (14.3.3) are *principal-value* integrals, that is, the limits as $\delta\to 0$ of the integrals over $V-V_\delta(\mathbf{r})$, where $V_\delta(\mathbf{r})$ is an excluded small sphere of radius δ centered about $\mathbf{r}$. See Appendix D and Refs. [1153,460,472,598] and [106–110] for the properties of such principal value integrals.

Similarly, the electric and magnetic fields (14.1.2) become:

$$\boxed{\begin{aligned}\boldsymbol{E}&=-\nabla\varphi-j\omega\boldsymbol{A}\\ \boldsymbol{H}&=\frac{1}{\mu}\nabla\times\boldsymbol{A}\end{aligned}} \tag{14.3.6}$$

With the help of the Lorenz condition the $\boldsymbol{E}$-field can be expressed completely in terms of the vector potential. Solving (14.3.5) for the scalar potential, $\varphi=-\nabla\cdot\boldsymbol{A}/j\omega\mu\epsilon$, and substituting in (14.3.6), we find

$$\boldsymbol{E}=\frac{1}{j\omega\mu\epsilon}\nabla(\nabla\cdot\boldsymbol{A})-j\omega\boldsymbol{A}=\frac{1}{j\omega\mu\epsilon}\big[\nabla(\nabla\cdot\boldsymbol{A})+k^2\boldsymbol{A}\big]$$

where we used $\omega^2\mu\epsilon=\omega^2/c^2=k^2$. To summarize, with $\boldsymbol{A}(\mathbf{r})$ computed from Eq. (14.3.1), the $\boldsymbol{E},\boldsymbol{H}$ fields are obtained from:

$$\boxed{\begin{aligned}\boldsymbol{E}&=\frac{1}{j\omega\mu\epsilon}\big[\nabla(\nabla\cdot\boldsymbol{A})+k^2\boldsymbol{A}\big]\\ \boldsymbol{H}&=\frac{1}{\mu}\nabla\times\boldsymbol{A}\end{aligned}} \tag{14.3.7}$$

An alternative way of expressing the electric field is:

$$\boxed{\boldsymbol{E}=\frac{1}{j\omega\mu\epsilon}\big[\nabla\times(\nabla\times\boldsymbol{A})-\mu\boldsymbol{J}\big]} \tag{14.3.8}$$

This is Ampère's law solved for $\boldsymbol{E}$. When applied to a *source-free* region of space, such as in the radiation zone, (14.3.8) simplifies into:

$$\boxed{\boldsymbol{E}=\frac{1}{j\omega\mu\epsilon}\nabla\times(\nabla\times\boldsymbol{A})} \tag{14.3.9}$$

The fields $\boldsymbol{E},\boldsymbol{H}$ can also be expressed directly in terms of the sources $\rho,\boldsymbol{J}$. Indeed, replacing the solutions (14.3.3) into Eqs. (14.3.6) or (14.3.7), we obtain:

$$\boxed{\begin{aligned}\boldsymbol{E}&=\int_V\Big[-j\omega\mu\boldsymbol{J}G+\frac{1}{\epsilon}\rho\nabla'G\Big]dV'=\frac{1}{j\omega\epsilon}\int_V\big[(\boldsymbol{J}\cdot\nabla')\nabla'G+k^2\boldsymbol{J}G\big]dV'\\ \boldsymbol{H}&=\int_V \boldsymbol{J}\times\nabla'G\,dV'\end{aligned}} \tag{14.3.10}$$

Here, $\rho,\boldsymbol{J}$ stand for $\rho(\mathbf{r}'),\boldsymbol{J}(\mathbf{r}')$. The gradient operator ∇ acts inside the integrands only on G and because that depends on the difference $\mathbf{r}-\mathbf{r}'$, we can replace the gradient with $\nabla G(\mathbf{r}-\mathbf{r}')=-\nabla'G(\mathbf{r}-\mathbf{r}')$. Also, we denoted $d^3\mathbf{r}'$ by dV'.

In obtaining (14.3.10), we had to interchange the operator ∇ and the integrals over V. When $\mathbf{r}$ is outside the volume V—as is the case for most of our applications—then, such interchanges are valid. When $\mathbf{r}$ lies within V, then, interchanging single ∇'s is still valid, as in the first expression for $\boldsymbol{E}$ and for $\boldsymbol{H}$. However, in interchanging double ∇'s, additional source terms arise. For example, using Eq. (D.8) of Appendix D, we find by interchanging the operator $\nabla\times\nabla\times$ with the integral for $\boldsymbol{A}$ in Eq. (14.3.8):

$$\boldsymbol{E}=\frac{1}{j\omega\epsilon}\big[\nabla\times\nabla\times\int_V \boldsymbol{J}G\,dV'-\boldsymbol{J}\big]=\frac{1}{j\omega\epsilon}\big[\frac{2}{3}\boldsymbol{J}+\mathrm{PV}\int_V\nabla\times\nabla\times(\boldsymbol{J}G)\,dV'-\boldsymbol{J}\big]$$

where “PV” stands for “principal value.” Because ∇ does not act on $\boldsymbol{J}(\mathbf{r}')$, we have:

$$\nabla\times\nabla\times(\boldsymbol{J}G)=\nabla\times(\nabla G\times\boldsymbol{J})=(\boldsymbol{J}\cdot\nabla)\nabla G-\boldsymbol{J}\nabla^2G=(\boldsymbol{J}\cdot\nabla')\nabla'G+k^2\boldsymbol{J}G$$

where in the last step, we replaced ∇ by $-\nabla'$ and $\nabla^2G=-k^2G$. It follows that:

$$\boldsymbol{E}=\frac{1}{j\omega\epsilon}\Big[\mathrm{PV}\int_V\big[(\boldsymbol{J}\cdot\nabla')\nabla'G+k^2\boldsymbol{J}G\big]dV'-\frac{1}{3}\boldsymbol{J}\Big],\quad(\mathbf{r}\text{ lies in }V)\tag{14.3.11}$$

In Sec. 17.10, we consider Eqs. (14.3.10) further in connection with Huygens’s principle and vector diffraction theory.

Next, we present three illustrative applications of the techniques discussed in this section: (a) Determining the fields of linear wire antennas, (b) The fields produced by electric and magnetic dipoles, and (c) the Ewald-Oseen extinction theorem and the microscopic origin of the refractive index. Then, we go on in Sec. 14.7 to discuss the simplification of the retarded potentials (14.3.3) for radiation problems.

14.4 Fields of a Linear Wire Antenna

Eqs. (14.3.7) simplify considerably in the special practical case of a linear wire antenna, that is, a *thin* cylindrical antenna. Figure 14.4.1 shows the geometry in the case of a z-directed antenna of finite length with a current $I(z')$ flowing on it.

The assumption that the radius of the wire is much smaller than its length means effectively that the current density $\boldsymbol{J}(\mathbf{r}')$ will be z-directed and confined to zero transverse dimensions, that is,

$$\boldsymbol{J}(\mathbf{r}')=\hat{\mathbf{z}}\,I(z')\,\delta(x')\,\delta(y')\qquad\text{(current on thin wire antenna)}\tag{14.4.1}$$

In the more realistic case of an antenna of finite radius a, the current density will be confined to flow on the cylindrical surface of the antenna, that is, at radial distance $\rho=a$. Assuming cylindrical symmetry, the current density will be:

$$\boldsymbol{J}(\mathbf{r}')=\hat{\mathbf{z}}\,I(z')\,\delta(\rho'-a)\,\frac{1}{2\pi a}\tag{14.4.2}$$

This case is discussed in more detail in Chap. 21. In both cases, integrating the current density over the transverse dimensions of the antenna gives the current:

$$\int\boldsymbol{J}(x',y',z')\,dx'dy'=\int\boldsymbol{J}(\rho',\phi',z')\,\rho'd\rho'd\phi'=\hat{\mathbf{z}}\,I(z')$$

Because of the cylindrical symmetry of the problem, the use of cylindrical coordinates is appropriate, especially in determining the fields *near* the antenna (cylindrical coordinates are reviewed in Sec. 14.8.) On the other hand, that the radiated fields at *far distances* from the antenna are best described in spherical coordinates. This is so because any finite current source appears as a point from far distances.

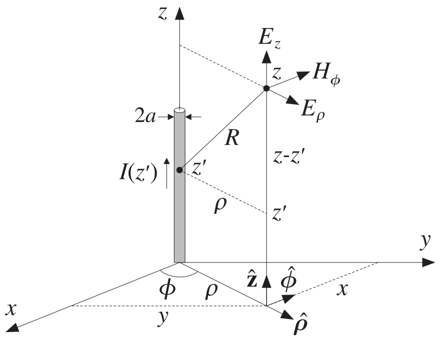


Fig. 14.4.1 Thin wire antenna.

Inserting Eq. (14.4.1) into Eq. (14.3.1), it follows that the vector potential will be z-directed and cylindrically symmetric. We have,

$$\begin{aligned}\boldsymbol{A}(\mathbf{r})&=\int_V\frac{\mu\boldsymbol{J}(\mathbf{r}')e^{-jkR}}{4\pi R}\,d^3\mathbf{r}'=\hat{\mathbf{z}}\,\frac{\mu}{4\pi}\int_V I(z')\,\delta(x')\,\delta(y')\,\frac{e^{-jkR}}{R}\,dx'dy'dz'\\&=\hat{\mathbf{z}}\,\frac{\mu}{4\pi}\int_L I(z')\,\frac{e^{-jkR}}{R}\,dz'\end{aligned}$$

where $R=|\mathbf{r}-\mathbf{r}'|=\sqrt{\rho^2+(z-z')^2}$, as shown in Fig. 14.4.1. The z'-integration is over the finite length of the antenna. Thus, $\boldsymbol{A}(\mathbf{r})=\hat{\mathbf{z}}\,A_z(\rho,z)$, with

$$\boxed{A_z(\rho,z)=\frac{\mu}{4\pi}\int_L I(z')\,\frac{e^{-jkR}}{R}\,dz'},\qquad R=\sqrt{\rho^2+(z-z')^2}\tag{14.4.3}$$

This is the solution of the z-component of the Helmholtz equation (14.3.2):

$$\nabla^2A_z+k^2A_z=-\mu I(z)\,\delta(x)\,\delta(y)$$

Because of the cylindrical symmetry, we can set $\partial/\partial\phi=0$. Therefore, the gradient and Laplacian operators are $\boldsymbol{\nabla}=\hat{\boldsymbol{\rho}}\,\partial_\rho+\hat{\mathbf{z}}\,\partial_z$ and $\nabla^2=\rho^{-1}\partial_\rho(\rho\partial_\rho)+\partial_z^2$. Thus, the Helmholtz equation can be written in the form:

$$\frac{1}{\rho}\partial_\rho(\rho\partial_\rho A_z)+\partial_z^2A_z+k^2A_z=-\mu I(z)\,\delta(x)\,\delta(y)$$

Away from the antenna, we obtain the homogeneous equation:

$$\frac{1}{\rho}\partial_\rho(\rho\partial_\rho A_z)+\partial_z^2A_z+k^2A_z=0\tag{14.4.4}$$

Noting that $\boldsymbol{\nabla}\cdot\boldsymbol{A}=\partial_zA_z$, we have from the Lorenz condition:

$$\boxed{\varphi = -\frac{1}{j\omega\mu\epsilon}\,\partial_z A_z} \qquad \text{(scalar potential of wire antenna)} \tag{14.4.5}$$

The z-component of the electric field is from Eq. (14.3.7):

$$j\omega\mu\epsilon E_z = \partial_z(\nabla\cdot\boldsymbol{A}) + k^2 A_z = \partial_z^2 A_z + k^2 A_z$$

and the radial component:

$$j\omega\mu\epsilon E_\rho = \partial_\rho(\nabla\cdot\boldsymbol{A}) = \partial_\rho\partial_z A_z$$

Using $\boldsymbol{B} = \nabla\times\boldsymbol{A} = (\hat{\boldsymbol{\rho}}\,\partial_\rho + \hat{\mathbf{z}}\,\partial_z)\times(\hat{\mathbf{z}}\,A_z) = (\hat{\boldsymbol{\rho}}\times\hat{\mathbf{z}})\,\partial_\rho A_z = -\hat{\boldsymbol{\phi}}\,\partial_\rho A_z$, it follows that the magnetic field has only a ϕ-component given by $B_\phi = -\partial_\rho A_z$. To summarize, the non-zero field components are all expressible in terms of A_z as follows:

$$\boxed{\begin{aligned} j\omega\mu\epsilon E_z &= \partial_z^2 A_z + k^2 A_z \\ j\omega\mu\epsilon E_\rho &= \partial_\rho\partial_z A_z \\ \mu H_\phi &= -\partial_\rho A_z \end{aligned}} \qquad \text{(fields of a wire antenna)} \tag{14.4.6}$$

Using Eq. (14.4.4), we may re-express E_z in the form:

$$j\omega\mu\epsilon E_z = -\frac{1}{\rho}\partial_\rho(\rho\partial_\rho A_z) = \mu\,\frac{1}{\rho}\partial_\rho(\rho H_\phi) \tag{14.4.7}$$

This is, of course, equivalent to the z-component of Ampère's law. In fact, an even more convenient way to construct the fields is to use the first of Eqs. (14.4.6) to construct E_z and then integrate Eq. (14.4.7) to get H_ϕ and then use the ρ-component of Ampère's law to get E_ρ. The resulting system of equations is:

$$\boxed{\begin{aligned} j\omega\mu\epsilon E_z &= \partial_z^2 A_z + k^2 A_z \\ \partial_\rho(\rho H_\phi) &= j\omega\epsilon\,\rho E_z \\ j\omega\epsilon E_\rho &= -\partial_z H_\phi \end{aligned}} \tag{14.4.8}$$

In Chap. 21, we use (14.4.6) to obtain the Hallén and Pocklington integral equations for determining the current $I(z)$ on a linear antenna, and solve them numerically. In Chap. 22, we use (14.4.8) under the assumption that the current $I(z)$ is sinusoidal to determine the near fields, and use them to compute the self and mutual impedances between linear antennas. The sinusoidal assumption for the current allows us to find E_z, and hence the rest of the fields, *without* having to find A_z first!

14.5 Fields of Electric and Magnetic Dipoles

Finding the fields produced by time-varying electric dipoles has been historically important and has served as a prototypical example for radiation problems.

We consider a point dipole located at the origin, in vacuum, with electric dipole moment $\mathbf{p}$. Assuming harmonic time dependence $e^{j\omega t}$, the corresponding polarization (dipole moment per unit volume) will be: $\boldsymbol{P}(\mathbf{r}) = \mathbf{p}\,\delta^{(3)}(\mathbf{r})$. We saw in Eq. (1.3.18) that the corresponding polarization current and charge densities are:

$$\boldsymbol{J} = \frac{\partial\boldsymbol{P}}{\partial t} = j\omega\boldsymbol{P}, \quad \rho = -\nabla\cdot\boldsymbol{P} \tag{14.5.1}$$

Therefore,

$$\boldsymbol{J}(\mathbf{r}) = j\omega\,\mathbf{p}\,\delta^{(3)}(\mathbf{r}), \quad \rho(\mathbf{r}) = -\mathbf{p}\cdot\nabla\delta^{(3)}(\mathbf{r}) \tag{14.5.2}$$

Because of the presence of the delta functions, the integrals in Eq. (14.3.3) can be done trivially, resulting in the vector and scalar potentials:

$$\begin{aligned} \boldsymbol{A}(\mathbf{r}) &= \mu_0\int j\omega\,\mathbf{p}\,\delta^{(3)}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dV' = j\omega\mu_0\,\mathbf{p}\,G(\mathbf{r}) \\ \varphi(\mathbf{r}) &= -\frac{1}{\epsilon_0}\int\big[\mathbf{p}\cdot\nabla'\delta^{(3)}(\mathbf{r}')\big]G(\mathbf{r}-\mathbf{r}')\,dV' = -\frac{1}{\epsilon_0}\,\mathbf{p}\cdot\nabla G(\mathbf{r}) \end{aligned} \tag{14.5.3}$$

where the integral for φ was done by parts. Alternatively, φ could have been determined from the Lorenz-gauge condition $\nabla\cdot\boldsymbol{A} + j\omega\mu_0\epsilon_0\varphi = 0$.

The $\boldsymbol{E},\boldsymbol{H}$ fields are computed from Eq. (14.3.6), or from (14.3.7), or away from the origin from (14.3.9). We find, where $k^2 = \omega^2/c_0^2 = \omega^2\mu_0\epsilon_0$:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= \frac{1}{\epsilon_0}\,\nabla\times\big[\nabla G(\mathbf{r})\times\mathbf{p}\big] = \frac{1}{\epsilon_0}\big[k^2\,\mathbf{p} + (\mathbf{p}\cdot\nabla)\nabla\big]G(\mathbf{r}) \\ \boldsymbol{H}(\mathbf{r}) &= j\omega\,\nabla G(\mathbf{r})\times\mathbf{p} \end{aligned}} \tag{14.5.4}$$

for $\mathbf{r}\neq 0$. The Green's function $G(\mathbf{r})$ and its gradient are:

$$G(\mathbf{r}) = \frac{e^{-jkr}}{4\pi r}, \quad \nabla G(\mathbf{r}) = -\hat{\mathbf{r}}\Big(jk + \frac{1}{r}\Big)G(\mathbf{r}) = -\hat{\mathbf{r}}\Big(jk + \frac{1}{r}\Big)\frac{e^{-jkr}}{4\pi r}$$

where $r = |\mathbf{r}|$ and $\hat{\mathbf{r}}$ is the radial unit vector $\hat{\mathbf{r}} = \mathbf{r}/r$. Inserting these into Eq. (14.5.4), we obtain the more explicit expressions:

$$\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= \frac{1}{\epsilon_0}\Big(jk + \frac{1}{r}\Big)\left[\frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{p}) - \mathbf{p}}{r}\right]G(\mathbf{r}) + \frac{k^2}{\epsilon_0}\,\hat{\mathbf{r}}\times(\mathbf{p}\times\hat{\mathbf{r}})\,G(\mathbf{r}) \\ \boldsymbol{H}(\mathbf{r}) &= j\omega\Big(jk + \frac{1}{r}\Big)(\mathbf{p}\times\hat{\mathbf{r}})\,G(\mathbf{r}) \end{aligned} \tag{14.5.5}$$

If the dipole is moved to location $\mathbf{r}_0$, so that $\boldsymbol{P}(\mathbf{r}) = \mathbf{p}\,\delta^{(3)}(\mathbf{r}-\mathbf{r}_0)$, then the fields are still given by Eqs. (14.5.4) and (14.5.5), with the replacement $G(\mathbf{r})\to G(\mathbf{R})$ and $\hat{\mathbf{r}}\to\hat{\mathbf{R}}$, where $\mathbf{R} = \mathbf{r}-\mathbf{r}_0$.

Eqs. (14.5.5) describe both the near fields and the radiated fields. The limit $\omega = 0$ (or $k = 0$) gives rise to the usual electrostatic dipole electric field, decreasing like $1/r^3$. On the other hand, as we discuss in Sec. 14.7, the radiated fields correspond to the terms decreasing like $1/r$. These are (with $\eta_0 = \sqrt{\mu_0/\epsilon_0}$):

$$
\begin{aligned}
\boldsymbol{E}_{\rm rad}(\mathbf{r}) &= \frac{k^2}{\epsilon_0}\,\hat{\mathbf{r}}\times(\mathbf{p}\times\hat{\mathbf{r}})\,G(\mathbf{r}) = \frac{k^2}{\epsilon_0}\,\hat{\mathbf{r}}\times(\mathbf{p}\times\hat{\mathbf{r}})\,\frac{e^{-jkr}}{4\pi r}\\
\boldsymbol{H}_{\rm rad}(\mathbf{r}) &= j\omega\, jk\,(\mathbf{p}\times\hat{\mathbf{r}})\,G(\mathbf{r}) = \frac{k^2}{\eta_0\epsilon_0}\,(\hat{\mathbf{r}}\times\mathbf{p})\,\frac{e^{-jkr}}{4\pi r}
\end{aligned}
\tag{14.5.6}
$$

They are related by $\eta_0\boldsymbol{H}_{\rm rad} = \hat{\mathbf{r}}\times\boldsymbol{E}_{\rm rad}$, which is a general relationship for radiation fields. The same expressions can also be obtained quickly from Eq. (14.5.4) by the substitution rule $\nabla\to -jk\hat{\mathbf{r}}$, discussed in Sec. 14.10.

The near-field, non-radiating, terms in (14.5.5) that drop faster than $1/r$ are important in the new area of *near-field optics* [494–514]. Nanometer-sized dielectric tips (constructed from a tapered fiber) act as tiny dipoles that can probe the evanescent fields from objects, resulting in a dramatic increase (by factors of ten) of the resolution of optical microscopy beyond the Rayleigh diffraction limit and down to atomic scales.

A magnetic dipole at the origin, with magnetic dipole moment $\mathbf{m}$, will be described by the magnetization vector $\boldsymbol{M} = \mathbf{m}\,\delta^{(3)}(\mathbf{r})$. According to Sec. 1.3, the corresponding magnetization current will be $\boldsymbol{J} = \nabla\times\boldsymbol{M} = \nabla\delta^{(3)}(\mathbf{r})\times\mathbf{m}$. Because $\nabla\cdot\boldsymbol{J} = 0$, there is no magnetic charge density, and hence, no scalar potential φ. The vector potential will be:

$$
\boldsymbol{A}(\mathbf{r}) = \mu_0\int\nabla\delta^{(3)}(\mathbf{r})\times\mathbf{m}\,G(\mathbf{r}-\mathbf{r}')\,dV' = \mu_0\,\nabla G(\mathbf{r})\times\mathbf{m}
\tag{14.5.7}
$$

It then follows from Eq. (14.3.6) that:

$$
\boxed{
\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= -j\omega\mu_0\nabla G(\mathbf{r})\times\mathbf{m}\\
\boldsymbol{H}(\mathbf{r}) &= \nabla\times\left[\nabla G(\mathbf{r})\times\mathbf{m}\right] = \left[k^2\,\mathbf{m} + (\mathbf{m}\cdot\nabla)\nabla\right]G(\mathbf{r})
\end{aligned}
}
\tag{14.5.8}
$$

which become explicitly,

$$
\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= j\omega\mu_0\big(jk+\frac{1}{r}\big)\,(\hat{\mathbf{r}}\times\mathbf{m})\,G(\mathbf{r})\\
\boldsymbol{H}(\mathbf{r}) &= \big(jk+\frac{1}{r}\big)\left[\frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{m})-\mathbf{m}}{r}\right]G(\mathbf{r}) + k^2\,\hat{\mathbf{r}}\times(\mathbf{m}\times\hat{\mathbf{r}})\,G(\mathbf{r})
\end{aligned}
\tag{14.5.9}
$$

The corresponding radiation fields are:

$$
\begin{aligned}
\boldsymbol{E}_{\rm rad}(\mathbf{r}) &= j\omega\mu_0\, jk\,(\hat{\mathbf{r}}\times\mathbf{m})\,G(\mathbf{r}) = \eta_0 k^2\,(\mathbf{m}\times\hat{\mathbf{r}})\,\frac{e^{-jkr}}{4\pi r}\\
\boldsymbol{H}_{\rm rad}(\mathbf{r}) &= k^2\,\hat{\mathbf{r}}\times(\mathbf{m}\times\hat{\mathbf{r}})\,G(\mathbf{r}) = k^2\,\hat{\mathbf{r}}\times(\mathbf{m}\times\hat{\mathbf{r}})\,\frac{e^{-jkr}}{4\pi r}
\end{aligned}
\tag{14.5.10}
$$

We note that the fields of the magnetic dipole are obtained from those of the electric dipole by the *duality* transformations $\boldsymbol{E}\to\boldsymbol{H}$, $\boldsymbol{H}\to-\boldsymbol{E}$, $\epsilon_0\to\mu_0$, $\mu_0\to\epsilon_0$, $\eta_0\to 1/\eta_0$, and $\mathbf{p}\to\mu_0\,\mathbf{m}$, that latter following by comparing the terms $\boldsymbol{P}$ and $\mu_0\boldsymbol{M}$ in the constitutive relations (1.3.16). Duality is discussed in more detail in Sec. 17.2.

The electric and magnetic dipoles are essentially equivalent to the linear and loop Hertzian dipole antennas, respectively, which are discussed in sections 16.2 and 16.8. Problem 14.4 establishes the usual results $\mathbf{p} = Q\,\mathbf{d}$ for a pair of charges $\pm Q$ separated by a distance $\mathbf{d}$, and $\mathbf{m} = \hat{\mathbf{z}}IS$ for a current loop of area S.

Example 14.5.1: We derive explicit expressions for the real-valued electric and magnetic fields of an oscillating z-directed dipole $\mathbf{p}(t) = p\,\hat{\mathbf{z}}\cos\omega t$. And also derive and plot the electric field lines at several time instants. This problem has an important history, having been considered first by Hertz in 1889 in a paper reprinted in [56].

Restoring the $e^{j\omega t}$ factor in Eq. (14.5.5) and taking real parts, we obtain the fields:

$$
\begin{aligned}
\boldsymbol{\mathcal{E}}(\mathbf{r}) &= p\big[k\sin(kr-\omega t) + \frac{\cos(kr-\omega t)}{r}\big]\,\frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\hat{\mathbf{z}})-\hat{\mathbf{z}}}{4\pi\epsilon_0 r^2} + \frac{pk^2\,\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\hat{\mathbf{r}})}{4\pi\epsilon_0 r}\cos(kr-\omega t)\\
\boldsymbol{\mathcal{H}}(\mathbf{r}) &= p\omega\big[-k\cos(kr-\omega t) + \frac{\sin(kr-\omega t)}{r}\big]\left[\frac{\hat{\mathbf{z}}\times\hat{\mathbf{r}}}{4\pi r}\right]
\end{aligned}
$$

In spherical coordinates, we have $\hat{\mathbf{z}} = \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta$. This gives $3\,\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\hat{\mathbf{z}})-\hat{\mathbf{z}} = 2\,\hat{\mathbf{r}}\cos\theta + \hat{\boldsymbol{\theta}}\sin\theta$, $\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\hat{\mathbf{r}}) = -\hat{\boldsymbol{\theta}}\sin\theta$, and $\hat{\mathbf{z}}\times\hat{\mathbf{r}} = \hat{\boldsymbol{\phi}}\sin\theta$. Therefore, the non-zero components of $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{H}}$ are $\mathcal{E}_r, \mathcal{E}_\phi$ and $\mathcal{H}_\phi$:

$$
\begin{aligned}
\mathcal{E}_r(\mathbf{r}) &= p\big[k\sin(kr-\omega t) + \frac{\cos(kr-\omega t)}{r}\big]\left[\frac{2\cos\theta}{4\pi\epsilon_0 r^2}\right]\\
\mathcal{E}_\theta(\mathbf{r}) &= p\big[k\sin(kr-\omega t) + \frac{\cos(kr-\omega t)}{r}\big]\left[\frac{\sin\theta}{4\pi\epsilon_0 r^2}\right] - \frac{pk^2\sin\theta}{4\pi\epsilon_0 r}\cos(kr-\omega t)\\
\mathcal{H}_\phi(\mathbf{r}) &= p\omega\big[-k\cos(kr-\omega t) + \frac{\sin(kr-\omega t)}{r}\big]\left[\frac{\sin\theta}{4\pi r}\right]
\end{aligned}
$$

By definition, the electric field is tangential to its field lines. A small displacement $d\mathbf{r}$ along the tangent to a line will be parallel to $\boldsymbol{\mathcal{E}}$ at that point. This implies that $d\mathbf{r}\times\boldsymbol{\mathcal{E}} = 0$, which can be used to determine the lines. Because of the azimuthal symmetry in the ϕ variable, we may look at the field lines that lie on the xz-plane (that is, $\phi = 0$). Then, we have:

$$
d\mathbf{r}\times\boldsymbol{\mathcal{E}} = (\hat{\mathbf{r}}\,dr + \hat{\boldsymbol{\theta}}\,r\,d\theta)\times(\hat{\mathbf{r}}\,\mathcal{E}_r + \hat{\boldsymbol{\theta}}\,\mathcal{E}_\theta) = \hat{\boldsymbol{\phi}}\,(dr\,\mathcal{E}_\theta - r\,d\theta\,\mathcal{E}_r) = 0 \quad\Rightarrow\quad \frac{dr}{d\theta} = \frac{r\,\mathcal{E}_r}{\mathcal{E}_\theta}
$$

This determines r as a function of θ, giving the polar representation of the line curve. To solve this equation, we rewrite the electric field in terms of the dimensionless variables $u = kr$ and $\delta = \omega t$, defining $E_0 = pk^3/4\pi\epsilon_0$:

$$
\begin{aligned}
\mathcal{E}_r &= E_0\,\frac{2\cos\theta}{u^2}\left[\sin(u-\delta) + \frac{\cos(u-\delta)}{u}\right]\\
\mathcal{E}_\theta &= -E_0\,\frac{\sin\theta}{u}\left[\cos(u-\delta) - \frac{\cos(u-\delta)}{u^2} - \frac{\sin(u-\delta)}{u}\right]
\end{aligned}
$$

We note that the factors within the square brackets are related by differentiation:

$$
\begin{aligned}
Q(u) &= \sin(u-\delta) + \frac{\cos(u-\delta)}{u}\\
Q'(u) &= \frac{dQ(u)}{du} = \cos(u-\delta) - \frac{\cos(u-\delta)}{u^2} - \frac{\sin(u-\delta)}{u}
\end{aligned}
$$

Therefore, the fields are:

$$
\mathcal{E}_r = E_0\,\frac{2\cos\theta}{u^2}\,Q(u)\,,\quad \mathcal{E}_\theta = -E_0\,\frac{\sin\theta}{u}\,Q'(u)
$$

It follows that the equation for the lines in the variable u will be:

$$\frac{du}{d\theta}=\frac{u\mathcal{E}_r}{\mathcal{E}_\theta}=-2\cot\theta\left[\frac{Q(u)}{Q'(u)}\right]\quad\Rightarrow\quad\frac{d}{d\theta}\big[\ln Q(u)\big]=-2\cot\theta=-\frac{d}{d\theta}\big[\ln\sin^2\theta\big]$$

which gives:

$$\frac{d}{d\theta}\ln\big[Q(u)\sin^2\theta\big]=0\quad\Rightarrow\quad Q(u)\sin^2\theta=C$$

where C is a constant. Thus, the electric field lines are given implicitly by:

$$\left[\sin(u-\delta)+\frac{\cos(u-\delta)}{u}\right]\sin^2\theta=\left[\sin(kr-\omega t)+\frac{\cos(kr-\omega t)}{kr}\right]\sin^2\theta=C$$

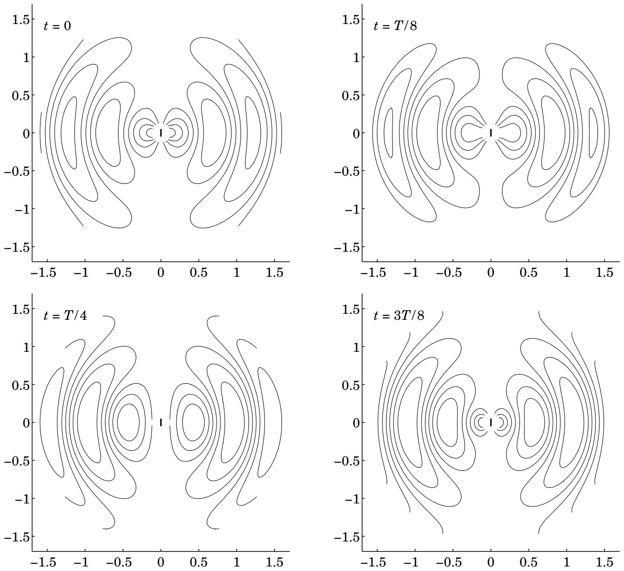


Fig. 14.5.1 Electric field lines of oscillating dipole at successive time instants.

Ideally, one should solve for r in terms of θ. Because this is not possible in closed form, we prefer to think of the lines as a contour plot at different values of the constant C. The resulting graphs are shown in Fig. 14.5.1. They were generated at the four time instants $t=0$, $T/8$, $T/4$, and $3T/8$, where T is the period of oscillation, $T=2\pi/\omega$. The x,z distances are in units of λ and extend to 1.5λ. The dipole is depicted as a tiny z-directed line at the origin. The following MATLAB code illustrates the generation of these plots:

```
rmin = 1/8; rmax = 1.6;                        % plot limits in wavelengths λ
Nr = 61; Nth = 61; N = 6;                      % meshpoints and number of contour levels
t = 1/8; d = 2*pi*t;                           % time instant t = T/8

[r,th] = meshgrid(linspace(rmin,rmax,Nr), linspace(0,pi,Nth));
```

```
u = 2*pi*r;                                    % r is in units of λ
z = r.*cos(th); x = r.*sin(th);                % cartesian coordinates in units of λ

C = (cos(u-d)./u + sin(u-d)) .* sin(th).^2;    % contour levels

contour([-x; x], [z; z], [C; C], N);           % right and left-reflected contours with N levels
```

We observe how the lines form closed loops originating at the dipole. The loops eventually escape the vicinity of the dipole and move outwards, pushing away the loops that are ahead of them. In this fashion, the field gets radiated away from its source. The MATLAB file `dipmovie.m` generates a movie of the evolving field lines lasting from $t=0$ to $t=8T$. □

14.6 Ewald-Oseen Extinction Theorem

The reflected and transmitted fields of a plane wave incident on a dielectric were determined in Chapters 5 and 7 by solving the wave equations in each medium and matching the solutions at the interface by imposing the boundary conditions.

Although this approach yields the correct solutions, it hides the physics. From the microscopic point of view, the dielectric consists of polarizable atoms or molecules, each of which is radiating *in vacuum* in response to the incident field and in response to the fields radiated by the other atoms. The total radiated field must combine with the incident field so as to generate the correct transmitted field. This is the essence of the Ewald-Oseen extinction theorem [458–493]. The word "extinction" refers to the cancellation of the incident field inside the dielectric.

Let $\boldsymbol{E}(\mathbf{r})$ be the incident field, $\boldsymbol{E}_{\rm rad}(\mathbf{r})$ the total radiated field, and $\boldsymbol{E}'(\mathbf{r})$ the transmitted field in the dielectric. Then, the theorem states that (for $\mathbf{r}$ inside the dielectric):

$$\boldsymbol{E}_{\rm rad}(\mathbf{r})=\boldsymbol{E}'(\mathbf{r})-\boldsymbol{E}(\mathbf{r})\quad\Rightarrow\quad\boxed{\boldsymbol{E}'(\mathbf{r})=\boldsymbol{E}(\mathbf{r})+\boldsymbol{E}_{\rm rad}(\mathbf{r})}\tag{14.6.1}$$

We will follow a simplified approach to the extinction theorem as in Refs. [479–493] and in particular [493]. We assume that the incident field is a uniform plane wave, with TE or TM polarization, incident obliquely on a planar dielectric interface, as shown in Fig. 14.6.1. The incident and transmitted fields will have the form:

$$\boldsymbol{E}(\mathbf{r})=\boldsymbol{E}_0\,e^{-j\,\boldsymbol{k}\cdot\mathbf{r}}\,,\quad\boldsymbol{E}'(\mathbf{r})=\boldsymbol{E}_0'\,e^{-j\,\boldsymbol{k}'\cdot\mathbf{r}}\tag{14.6.2}$$

The expected relationships between the transmitted and incident waves were summarized in Eqs. (7.7.1)–(7.7.5). We will derive the same results from the present approach. The incident wave vector is $\boldsymbol{k}=k_x\,\hat{\mathbf{z}}+k_z\,\hat{\mathbf{z}}$ with $k=\omega/c_0=\omega\sqrt{\epsilon_0\mu_0}$, and satisfies $\boldsymbol{k}\cdot\boldsymbol{E}_0=0$. For the transmitted wave, we will find that $\boldsymbol{k}'=k_x\,\hat{\mathbf{z}}+k_z'\,\hat{\mathbf{z}}$ satisfies $\boldsymbol{k}'\cdot\boldsymbol{E}_0'=0$ and $k'=\omega/c=\omega\sqrt{\epsilon\mu_0}=kn$, so that $c=c_0/n$, where n is the refractive index of the dielectric, $n=\sqrt{\epsilon/\epsilon_0}$.

The radiated field is given by Eq. (14.3.10), where $\boldsymbol{J}$ is the current due to the polarization $\boldsymbol{P}$, that is, $\boldsymbol{J}=\dot{\boldsymbol{P}}=j\omega\boldsymbol{P}$. Although there is no volume polarization charge density,[†] there may be a surface polarization density $\rho_s=\hat{\mathbf{n}}\cdot\boldsymbol{P}$ on the planar dielectric interface. Because $\hat{\mathbf{n}}=-\hat{\mathbf{z}}$, we will have $\rho_s=-\hat{\mathbf{z}}\cdot\boldsymbol{P}=-P_z$. Such density is present only in the TM

[†] $\rho=-\nabla\cdot\boldsymbol{P}$ vanishes for the type of plane-wave solutions that we consider here.

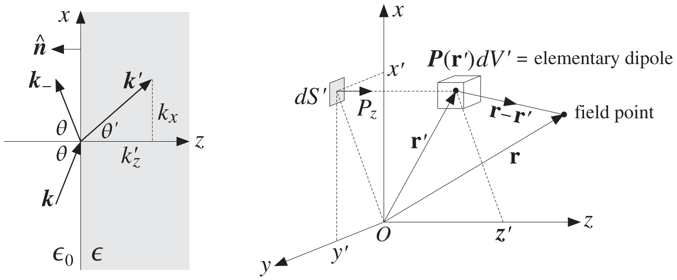


Fig. 14.6.1 Elementary dipole at $\mathbf{r}'$ contributes to the local field at $\mathbf{r}$.

case [493]. The corresponding volume term in Eq. (14.3.10) will collapse into a surface integral. Thus, the field generated by the densities $\boldsymbol{J}, \rho_s$ will be:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = -j\omega\mu_0 \int_V \boldsymbol{J}(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dV' + \frac{1}{\epsilon_0}\int_S \rho_s(\mathbf{r}')\, \nabla' G(\mathbf{r}-\mathbf{r}')\, dS'$$

where $G(\mathbf{r}) = e^{-jkr}/4\pi r$ is the vacuum Green's function having $k = \omega/c_0$, and V is the right half-space $z \ge 0$, and S, the xy-plane. Replacing $\boldsymbol{J}, \rho_s$ in terms of the polarization and writing $\nabla' G = -\nabla G$, and moving ∇ outside the surface integral, we have:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = \omega^2\mu_0 \int_V \boldsymbol{P}(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dV' + \frac{1}{\epsilon_0}\nabla \int_S P_z(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dS' \tag{14.6.3}$$

We assume that the polarization $\boldsymbol{P}(\mathbf{r}')$ is induced by the total field inside the dielectric, that is, we set $\boldsymbol{P}(\mathbf{r}') = \epsilon_0\chi\,\boldsymbol{E}'(\mathbf{r}')$, where χ is the electric susceptibility. Setting $k^2 = \omega^2\mu_0\epsilon_0$, Eq. (14.6.3) becomes:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = k^2\chi \int_V \boldsymbol{E}'(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dV' + \chi\nabla \int_S E'_z(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dS' \tag{14.6.4}$$

Evaluated at points $\mathbf{r}$ on the left of the interface ($z < 0$), $\boldsymbol{E}_{\text{rad}}(\mathbf{r})$ should generate the *reflected field*. Evaluated within the dielectric ($z \ge 0$), it should give Eq. (14.6.1), resulting in the self-consistency condition:

$$k^2\chi \int_V \boldsymbol{E}'(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dV' + \chi\nabla \int_S E'_z(\mathbf{r}')\, G(\mathbf{r}-\mathbf{r}')\, dS' = \boldsymbol{E}'(\mathbf{r}) - \boldsymbol{E}(\mathbf{r}) \tag{14.6.5}$$

Inserting Eq. (14.6.2), we obtain the condition:

$$k^2\chi\,\boldsymbol{E}'_0 \int_V e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dV' + \chi E'_{z0}\nabla \int_S e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dS' = \boldsymbol{E}'_0\, e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\, e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

The vector $\boldsymbol{k}' = k'_x\hat{\mathbf{x}} + k'_z\hat{\mathbf{z}}$ may be assumed to have $k'_x = k_x$, which is equivalent to Snel's law. This follows easily from the phase matching of the $e^{jk_x x}$ factors in the above equation. Then, the integrals over S and V can be done easily using Eqs. (D.14) and (D.16) of Appendix D, with (D.14) being evaluated at $z' = 0$ and $z \ge 0$:

$$\begin{aligned}
&\int_V e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dV' = \frac{e^{-j\boldsymbol{k}'\cdot\mathbf{r}}}{k'^2-k^2} - \frac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z(k'_z-k_z)} \\
&\int_S e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dS' = \frac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2jk_z} \;\Rightarrow\; \nabla\int_S e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dS' = -\frac{\boldsymbol{k}\, e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z}
\end{aligned} \tag{14.6.6}$$

The self-consistency condition reads now:

$$k^2\chi\,\boldsymbol{E}'_0\left[\frac{e^{-j\boldsymbol{k}'\cdot\mathbf{r}}}{k'^2-k^2} - \frac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z(k'_z-k_z)}\right] - \chi E'_{z0}\,\frac{\boldsymbol{k}\, e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z} = \boldsymbol{E}'_0\, e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\, e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

Equating the coefficients of like exponentials, we obtain the two conditions:

$$\frac{k^2\chi}{k'^2-k^2}\,\boldsymbol{E}'_0 = \boldsymbol{E}'_0 \quad\Rightarrow\quad \frac{k^2\chi}{k'^2-k^2} = 1 \quad\Rightarrow\quad k'^2 = k^2(1+\chi) = k^2n^2 \tag{14.6.7}$$

$$\frac{k^2\chi}{2k_z(k'_z-k_z)}\,\boldsymbol{E}'_0 + \frac{\chi\,\boldsymbol{k}}{2k_z}\,E'_{z0} = \boldsymbol{E}_0 \tag{14.6.8}$$

The first condition implies that $k' = kn$, where $n = \sqrt{1+\chi} = \sqrt{\epsilon/\epsilon_0}$. Thus, the phase velocity within the dielectric is $c = c_0/n$. Replacing $\chi = (k'^2-k^2)/k^2 = (k'^2_z-k^2_z)/k^2$, we may rewrite Eq. (14.6.8) as:

$$\frac{k'^2_z-k^2_z}{2k_z(k'_z-k_z)}\,\boldsymbol{E}'_0 + \frac{(k'^2_z-k^2_z)\,\boldsymbol{k}}{2k_z k^2}\,E'_{z0} = \boldsymbol{E}_0\,, \quad\text{or,}$$

$$\boldsymbol{E}'_0 + \frac{\boldsymbol{k}}{k^2}(k'_z-k_z)E'_{z0} = \frac{2k_z}{k'_z+k_z}\,\boldsymbol{E}_0 \tag{14.6.9}$$

This implies immediately the transversality condition for the transmitted field, that is, $\boldsymbol{k}'\cdot\boldsymbol{E}'_0 = 0$. Indeed, using $\boldsymbol{k}\cdot\boldsymbol{E}_0 = 0$ for the incident field, we find:

$$\boldsymbol{k}\cdot\boldsymbol{E}'_0 + \frac{\boldsymbol{k}\cdot\boldsymbol{k}}{k^2}(k'_z-k_z)E'_{z0} = \frac{2k_z}{k'_z+k_z}\,\boldsymbol{k}\cdot\boldsymbol{E}_0 = 0 \quad\Rightarrow\quad \boldsymbol{k}\cdot\boldsymbol{E}'_0 + (k'_z-k_z)E'_{z0} = 0$$

or, explicitly, $k_xE'_{x0} + k_zE'_{z0} + (k'_z-k_z)E'_{z0} = k_xE'_{x0} + k'_zE'_{z0} = \boldsymbol{k}'\cdot\boldsymbol{E}'_0 = 0$. Replacing $(k'_z-k_z)E'_{z0} = -\boldsymbol{k}\cdot\boldsymbol{E}'_0$ in Eq. (14.6.9) and using the BAC-CAB rule, we obtain:

$$\boldsymbol{E}'_0 - \frac{\boldsymbol{k}}{k^2}(\boldsymbol{k}\cdot\boldsymbol{E}'_0) = \frac{2k_z}{k'_z+k_z}\,\boldsymbol{E}_0 \quad\Rightarrow\quad \boxed{\frac{\boldsymbol{k}\times(\boldsymbol{E}'_0\times\boldsymbol{k})}{k^2} = \frac{2k_z}{k'_z+k_z}\,\boldsymbol{E}_0} \tag{14.6.10}$$

It can be shown that Eq. (14.6.10) is *equivalent* to the transmission coefficient results summarized in Eqs. (7.7.1)–(7.7.5), for both the TE and TM cases (see also Problem 7.6 and the identities in Problem 7.5.) The transmitted magnetic field $\boldsymbol{H}'(\mathbf{r}) = \boldsymbol{H}'_0\, e^{-j\boldsymbol{k}'\cdot\mathbf{r}}$ may be found from Faraday's law $\nabla\times\boldsymbol{E}' = -j\omega\mu_0\boldsymbol{H}'$, which reads $\omega\mu_0\boldsymbol{H}'_0 = \boldsymbol{k}'\times\boldsymbol{E}'_0$.

Next, we look at the reflected field. For points $\mathbf{r}$ lying to the left of the interface ($z \le 0$), the evaluation of the integrals (14.6.6) gives according to Eqs. (D.14) and (D.16), where (D.14) is evaluated at $z' = 0$ and $z \le 0$:

$$\begin{aligned}
&\int_V e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dV' = -\frac{e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2k_z(k'_z+k_z)} \\
&\int_S e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dS' = \frac{e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2jk_z} \;\Rightarrow\; \nabla\int_S e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\, G(\mathbf{r}-\mathbf{r}')\, dS' = -\frac{\boldsymbol{k}_-\, e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2k_z}
\end{aligned}$$

where $\boldsymbol{k}_-$ denotes the reflected wave vector, $\boldsymbol{k}_- = k_x\hat{\mathbf{x}} - k_z\hat{\mathbf{z}}$. It follows that the total radiated field will be:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = k^2\chi\,\boldsymbol{E}'_0\left[-\frac{e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2k_z(k'_z+k_z)}\right] - \frac{\boldsymbol{k}_-\,\chi E'_{z0}}{2k_z}\, e^{-j\boldsymbol{k}_-\cdot\mathbf{r}} = \boldsymbol{E}_{-0}\, e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}$$

where the overall coefficient $\boldsymbol{E}_{-0}$ can be written in the form:

$$\boldsymbol{E}_{-0} = -\frac{k^2\chi}{2k_z(k_z'+k_z)}\boldsymbol{E}_0' - \frac{\boldsymbol{k}_-\chi E_{z0}'}{2k_z} = \frac{k_z-k_z'}{2k_z}\left[\boldsymbol{E}_0' + \frac{\boldsymbol{k}_-(k_z'+k_z)E_{z0}'}{k^2}\right]$$

where we set $\chi = (k_z'^2 - k_z^2)/k^2$. Noting the identity $\boldsymbol{k}_-\cdot\boldsymbol{E}_0' + (k_z'+k_z)E_{z0}' = \boldsymbol{k}'\cdot\boldsymbol{E}_0' = 0$ and $\boldsymbol{k}_-\cdot\boldsymbol{k}_- = k^2$, we finally find:

$$\boldsymbol{E}_{-0} = \frac{k_z-k_z'}{2k_z}\left[\boldsymbol{E}_0' - \frac{\boldsymbol{k}_-(\boldsymbol{k}_-\cdot\boldsymbol{E}_0')}{k^2}\right] \quad\Rightarrow\quad \boxed{\frac{\boldsymbol{k}_-\times(\boldsymbol{E}_0'\times\boldsymbol{k}_-)}{k^2} = \frac{2k_z}{k_z-k_z'}\boldsymbol{E}_{-0}} \tag{14.6.11}$$

It can be verified that (14.6.11) is equivalent to the reflected fields as given by Eqs. (7.7.1)–(7.7.5) for the TE and TM cases. We note also that $\boldsymbol{k}_-\cdot\boldsymbol{E}_{-0} = 0$.

The conventional boundary conditions are a *consequence* of this approach. For example, Eqs. (14.6.10) and (14.6.11) imply the continuity of the tangential components of the $\boldsymbol{E}$-field. Indeed, we find by adding:

$$\boldsymbol{E}_0 + \boldsymbol{E}_{-0} = \boldsymbol{E}_0' + \frac{\chi E_{z0}'}{2k_z}(\boldsymbol{k}-\boldsymbol{k}_-) = \boldsymbol{E}_0' + \chi\,\hat{\mathbf{z}}\,E_{z0}'$$

which implies that $\hat{\mathbf{z}}\times(\boldsymbol{E}_0+\boldsymbol{E}_{-0}) = \hat{\mathbf{z}}\times\boldsymbol{E}_0'$.

In summary, the radiated fields from the polarizable atoms cause the cancellation of the incident vacuum field throughout the dielectric and conspire to generate the correct transmitted field that has phase velocity $c = c_0/n$. The reflected wave does not originate just at the interface but rather it is the field radiated backwards by the atoms within the entire body of the dielectric.

Next, we discuss another simplified approach based on radiating dipoles [484]. It has the additional advantage that it leads to the Lorentz-Lorenz or Clausius-Mossotti relationship between refractive index and polarizability. General proofs of the extinction theorem may be found in [458–478] and [598].

The dielectric is viewed as a collection of dipoles $\mathbf{p}_i$ at locations $\mathbf{r}_i$. The dipole moments are assumed to be induced by a local (or effective) electric field $\boldsymbol{E}_{\text{loc}}(\mathbf{r})$ through $\mathbf{p}_i = \alpha\epsilon_0\boldsymbol{E}_{\text{loc}}(\mathbf{r}_i)$, where α is the polarizability.[†] The field radiated by the jth dipole $\mathbf{p}_j$ is given by Eq. (14.5.4), where $G(\mathbf{r})$ is the vacuum Green's function:

$$\boldsymbol{E}_j(\mathbf{r}) = \frac{1}{\epsilon_0}\nabla\times\nabla\times\left[\mathbf{p}_j\,G(\mathbf{r}-\mathbf{r}_j)\right]$$

The field at the location of the ith dipole due to all the other dipoles will be:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}_i) = \sum_{j\neq i}\boldsymbol{E}_j(\mathbf{r}_i) = \frac{1}{\epsilon_0}\sum_{j\neq i}\nabla_i\times\nabla_i\times\left[\mathbf{p}_j\,G(\mathbf{r}_i-\mathbf{r}_j)\right] \tag{14.6.12}$$

where ∇_i is with respect to $\mathbf{r}_i$. Passing to a continuous description, we assume N dipoles per unit volume, so that the polarization density will be $\boldsymbol{P}(\mathbf{r}') = N\,\mathbf{p}(\mathbf{r}') = N\alpha\epsilon_0\boldsymbol{E}_{\text{loc}}(\mathbf{r}')$. Then, Eq. (14.6.12) is replaced by the (principal-value) integral:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = \frac{1}{\epsilon_0}\int_V\Big[\nabla\times\nabla\times\left[\boldsymbol{P}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\right]\Big]_{\mathbf{r}'\neq\mathbf{r}}dV' \tag{14.6.13}$$

[†] Normally, the polarizability is defined as the quantity $\alpha' = \alpha\epsilon_0$.

Using Eq. (D.7) of Appendix D, we rewrite:

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = \frac{1}{\epsilon_0}\nabla\times\nabla\times\int_V\boldsymbol{P}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dV' - \frac{2}{3\epsilon_0}\boldsymbol{P}(\mathbf{r}) \tag{14.6.14}$$

and in terms of the local field ($N\alpha$ is dimensionless):

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = N\alpha\,\nabla\times\nabla\times\int_V\boldsymbol{E}_{\text{loc}}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dV' - \frac{2}{3}N\alpha\,\boldsymbol{E}_{\text{loc}}(\mathbf{r}) \tag{14.6.15}$$

According to the Ewald-Oseen extinction requirement, the radiated field must cancel the incident field $\boldsymbol{E}(\mathbf{r})$ while generating the local field $\boldsymbol{E}_{\text{loc}}(\mathbf{r})$, that is, $\boldsymbol{E}_{\text{rad}}(\mathbf{r}) = \boldsymbol{E}_{\text{loc}}(\mathbf{r}) - \boldsymbol{E}(\mathbf{r})$. This leads to the self-consistency condition:

$$N\alpha\,\nabla\times\nabla\times\int_V\boldsymbol{E}_{\text{loc}}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dV' - \frac{2}{3}N\alpha\,\boldsymbol{E}_{\text{loc}}(\mathbf{r}) = \boldsymbol{E}_{\text{loc}}(\mathbf{r}) - \boldsymbol{E}(\mathbf{r}) \tag{14.6.16}$$

Assuming a plane-wave solution $\boldsymbol{E}_{\text{loc}}(\mathbf{r}) = \boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}}$, we obtain:

$$N\alpha\,\nabla\times\nabla\times\boldsymbol{E}_1'\int_V e^{-j\boldsymbol{k}'\cdot\mathbf{r}}\,G(\mathbf{r}-\mathbf{r}')\,dV' - \frac{2}{3}N\alpha\,\boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} = \boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\,e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

For $\mathbf{r}$ within the dielectric, we find as before:

$$N\alpha\,\nabla\times\nabla\times\boldsymbol{E}_1'\left[\frac{e^{-j\boldsymbol{k}'\cdot\mathbf{r}}}{k'^2-k^2} - \frac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z(k_z'-k_z)}\right] - \frac{2}{3}N\alpha\,\boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} = \boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\,e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

$$N\alpha\,\nabla\times\nabla\times\boldsymbol{E}_1'\left[\frac{e^{-j\boldsymbol{k}'\cdot\mathbf{r}}}{k'^2-k^2} - \frac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z(k_z'-k_z)}\right] = \left(1+\frac{2}{3}N\alpha\right)\boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\,e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

Performing the ∇ operations, we have:

$$N\alpha\left[\frac{\boldsymbol{k}'\times(\boldsymbol{E}_1'\times\boldsymbol{k}')}{k'^2-k^2}e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \frac{\boldsymbol{k}\times(\boldsymbol{E}_1'\times\boldsymbol{k})}{2k_z(k_z'-k_z)}e^{-j\boldsymbol{k}\cdot\mathbf{r}}\right] = \left(1+\frac{2}{3}N\alpha\right)\boldsymbol{E}_1'\,e^{-j\boldsymbol{k}'\cdot\mathbf{r}} - \boldsymbol{E}_0\,e^{-j\boldsymbol{k}\cdot\mathbf{r}}$$

Equating the coefficients of the exponentials, we obtain the two conditions:

$$N\alpha\,\frac{\boldsymbol{k}'\times(\boldsymbol{E}_1'\times\boldsymbol{k}')}{k'^2-k^2} = \left(1+\frac{2}{3}N\alpha\right)\boldsymbol{E}_1' \tag{14.6.17}$$

$$N\alpha\,\frac{\boldsymbol{k}\times(\boldsymbol{E}_1'\times\boldsymbol{k})}{2k_z(k_z'-k_z)} = \boldsymbol{E}_0 \tag{14.6.18}$$

The first condition implies immediately that $\boldsymbol{k}'\cdot\boldsymbol{E}_1' = 0$, therefore, using the BAC-CAB rule, the condition reads:

$$\frac{N\alpha\,k'^2}{k'^2-k^2}\boldsymbol{E}_1' = \left(1+\frac{2}{3}N\alpha\right)\boldsymbol{E}_1' \quad\Rightarrow\quad \frac{N\alpha\,k'^2}{k'^2-k^2} = 1+\frac{2}{3}N\alpha \tag{14.6.19}$$

Setting $k' = kn$, Eq. (14.6.19) implies the *Lorentz-Lorenz formula*:

$$\frac{N\alpha\,n^2}{n^2-1} = 1+\frac{2}{3}N\alpha \quad\Rightarrow\quad \boxed{\frac{n^2-1}{n^2+2} = \frac{1}{3}N\alpha} \tag{14.6.20}$$

We must distinguish between the local field $\boldsymbol{E}_{\text{loc}}(\mathbf{r})$ and the measured or observed field $\boldsymbol{E}'(\mathbf{r})$, the latter being a "screened" version of the former. To find their relationship, we define the susceptibility by $\chi = n^2 - 1$ and require that the polarization $\boldsymbol{P}(\mathbf{r})$ be related to the observed field by the usual relationship $\boldsymbol{P} = \epsilon_0 \chi \boldsymbol{E}'$. Using the Lorentz-Lorenz formula and $\boldsymbol{P} = N\alpha\epsilon_0 \boldsymbol{E}_{\text{loc}}$, we find the well-known relationship [598]:

$$\boxed{\boldsymbol{E}_{\text{loc}} = \boldsymbol{E}' + \frac{\boldsymbol{P}}{3\epsilon_0}} \tag{14.6.21}$$

From $N\alpha\boldsymbol{E}_{\text{loc}} = \boldsymbol{P}/\epsilon_0 = \chi\boldsymbol{E}'$, we have $N\alpha\boldsymbol{E}_1' = \chi\boldsymbol{E}_0'$. Then, the second condition (14.6.18) may be expressed in terms of $\boldsymbol{E}_0'$:

$$\frac{\chi\,\boldsymbol{k}\times(\boldsymbol{E}_0'\times\boldsymbol{k})}{2k_z(k_z'-k_z)} = \boldsymbol{E}_0 \quad\Rightarrow\quad \frac{\boldsymbol{k}\times(\boldsymbol{E}_0'\times\boldsymbol{k})}{k^2} = \frac{2k_z}{k_z'+k_z}\,\boldsymbol{E}_0 \tag{14.6.22}$$

which is identical to Eq. (14.6.10). Thus, the self-consistent solution for $\boldsymbol{E}'(\mathbf{r})$ is identical to that found previously.

Finally, we obtain the reflected field by evaluating Eq. (14.6.13) at points $\mathbf{r}$ to the left of the interface. In this case, there is no $2\boldsymbol{P}/3\epsilon_0$ term in (14.6.14) and we have:

$$\begin{aligned}
\boldsymbol{E}_{\text{rad}}(\mathbf{r}) &= N\alpha\,\nabla\times\nabla\times\int_V \boldsymbol{E}_{\text{loc}}(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV' = \chi\,\nabla\times\nabla\times\int_V \boldsymbol{E}'(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV' \\
&= \chi\,\nabla\times\nabla\times\boldsymbol{E}_0'\int_V e^{-j\boldsymbol{k}'\cdot\mathbf{r}}G(\mathbf{r}-\mathbf{r}')\,dV' = \chi\,\nabla\times\nabla\times\boldsymbol{E}_0'\left[-\frac{e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2k_z(k_z'+k_z)}\right] \\
&= -\frac{\chi\,\boldsymbol{k}_-\times(\boldsymbol{E}_0'\times\boldsymbol{k}_-)}{2k_z(k_z'+k_z)}\,e^{-j\boldsymbol{k}_-\cdot\mathbf{r}} = \frac{k_z-k_z'}{2k_z}\,\frac{\boldsymbol{k}_-\times(\boldsymbol{E}_0'\times\boldsymbol{k}_-)}{k^2}\,e^{-j\boldsymbol{k}_-\cdot\mathbf{r}} = \boldsymbol{E}_{-0}\,e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}
\end{aligned}$$

which agrees with Eq. (14.6.11).

14.7 Radiation Fields

The retarded solutions (14.3.3) for the potentials are quite general and apply to any current and charge distribution. Here, we begin making a number of approximations that are relevant for radiation problems. We are interested in fields that have radiated away from their current sources and are capable of carrying power to large distances from the sources.

The *far-field approximation* assumes that the field point $\mathbf{r}$ is very far from the current source. Here, "far" means much farther than the typical spatial extent of the current distribution, that is, $r \gg r'$. Because r' varies only over the current source we can state this condition as $r \gg l$, where l is the typical extent of the current distribution (for example, for a linear antenna, l is its length.) Fig. 14.7.1 shows this approximation.

As shown in Fig. 14.7.1, at far distances the sides PP′ and PQ of the triangle PQP′ are almost equal. But the side PQ is the difference OP−OQ. Thus, $R \simeq r - \hat{\mathbf{r}}\cdot\mathbf{r}' = r - r'\cos\psi$, where ψ is the angle between the vectors $\mathbf{r}$ and $\mathbf{r}'$.

A better approximation may be obtained with the help of the small-x Taylor series expansion $\sqrt{1+x} \simeq 1 + x/2 - x^2/8$. Expanding R in powers of r'/r, and keeping terms up to second order, we obtain:

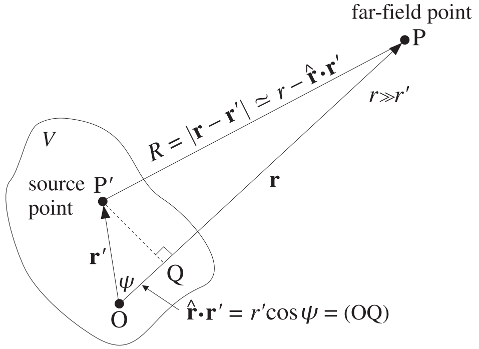

far-field point
P
$R = |\mathbf{r}-\mathbf{r}'| \simeq r - \hat{\mathbf{r}}\cdot\mathbf{r}'$
$r \gg r'$
V
source point
P′
$\mathbf{r}$
$\mathbf{r}'$
ψ
Q
O
$\hat{\mathbf{r}}\cdot\mathbf{r}' = r'\cos\psi = (\text{OQ})$

Fig. 14.7.1 Far-field approximation.

$$\begin{aligned}
R = |\mathbf{r}-\mathbf{r}'| &= \sqrt{r^2 - 2rr'\cos\psi + r'^2} = r\sqrt{1 - 2\frac{r'}{r}\cos\psi + \frac{r'^2}{r^2}} \\
&\simeq r\left(1 - \frac{r'}{r}\cos\psi + \frac{r'^2}{2r^2} - \frac{1}{8}\big(-2\frac{r'}{r}\cos\psi + \frac{r'^2}{r^2}\big)^2\right) \\
&\simeq r\big(1 - \frac{r'}{r}\cos\psi + \frac{r'^2}{2r^2} - \frac{r'^2}{2r^2}\cos^2\psi)\big)
\end{aligned}$$

or, combining the last two terms:

$$R = r - r'\cos\psi + \frac{r'^2}{2r}\sin^2\psi\,, \qquad \text{for}\quad r \gg r' \tag{14.7.2}$$

Thus, the first-order approximation is $R = r - r'\cos\psi = r - \hat{\mathbf{r}}\cdot\mathbf{r}'$. Using this approximation in the integrands of Eqs. (14.3.1), we have:

$$\varphi(\mathbf{r}) \simeq \int_V \frac{\rho(\mathbf{r}')e^{-jk(r-\hat{\mathbf{r}}\cdot\mathbf{r}')}}{4\pi\epsilon(r-\hat{\mathbf{r}}\cdot\mathbf{r}')}\,d^3\mathbf{r}'$$

Replacing $R = r - \hat{\mathbf{r}}\cdot\mathbf{r}' \simeq r$ in the denominator, but not in the exponent, we obtain the far-field approximation to the solution:

$$\varphi(\mathbf{r}) = \frac{e^{-jkr}}{4\pi\epsilon r}\int_V \rho(\mathbf{r}')e^{jk\hat{\mathbf{r}}\cdot\mathbf{r}'}\,d^3\mathbf{r}'$$

Because R is approximated differently in the denominator and the exponent, it might be argued that we are not making a consistent approximation. Indeed, for multipole expansions, it is not correct to ignore the $\hat{\mathbf{r}}\cdot\mathbf{r}'$ term from the denominator. However, the procedure is correct for *radiation problems*, and generates those terms that correspond to propagating waves.

What about the second-order approximation terms? We have dropped them from both the exponent and the denominator. Because in the exponent they are multiplied by k, in order to justify dropping them, we must require in addition to $r \gg r'$ that $kr'^2/r \ll 1$, or in terms of the wavelength: $r \gg 2\pi r'^2/\lambda$. Replacing $2r'$ by the typical size l of the current source,[†] we have $r \gg \pi l^2/2\lambda$. By convention [92], we replace this with $r \gg 2l^2/\lambda$. Thus, we may state the far-field conditions as:

$$r \gg l \qquad \text{and} \qquad r \gg \frac{2l^2}{\lambda} \qquad \text{(far-field conditions)} \tag{14.7.3}$$

These conditions define the so-called far-field or Fraunhofer radiation region. They are easily satisfied for many practical antennas (such as the half-wave dipole) because l is typically of the same order of magnitude as λ, in which case the second condition is essentially equivalent to the first. This happens also when $l > \lambda$. When $l \ll \lambda$, the first condition implies the second.

The distance $r = 2l^2/\lambda$ is by convention [92] the dividing line between the far-field (Fraunhofer) region, and the near-field (Fresnel) region, as shown in Fig. 14.7.2. The far-field region is characterized by the property that the *angular* distribution of radiation is independent of the distance r.

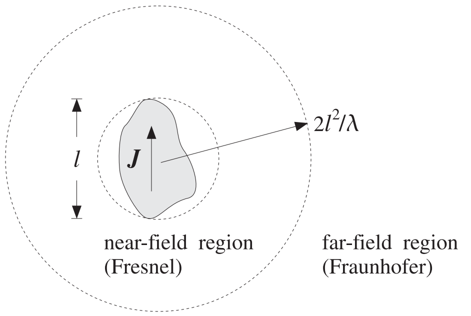


Fig. 14.7.2 Far-field and near-field radiation zones.

Can the first-order term $k\hat{\mathbf{r}}\cdot\mathbf{r}'$ also be ignored from the exponent? This would require that $kr' \ll 1$, or that $r' \ll \lambda$. Thus, it can be ignored for electrically "short" antennas, that is, $l \ll \lambda$, or equivalently in the long wavelength or low-frequency limit. The Hertzian dipole is such an antenna example.

Defining the wavenumber vector $\boldsymbol{k}$ to be in the direction of the field vector $\mathbf{r}$ and having magnitude k, that is, $\boldsymbol{k} = k\hat{\mathbf{r}}$, we may summarize the *far-field approximation* to the retarded single-frequency potentials as follows:

$$\boxed{\begin{aligned} \varphi(\mathbf{r}) &= \frac{e^{-jkr}}{4\pi\epsilon r}\int_V \rho(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' \\ \boldsymbol{A}(\mathbf{r}) &= \frac{\mu e^{-jkr}}{4\pi r}\int_V \boldsymbol{J}(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' \end{aligned}}\,, \qquad \boxed{\boldsymbol{k} = k\hat{\mathbf{r}}} \tag{14.7.4}$$

[†] We envision a sphere of diameter $2r' = l$ enclosing the antenna structure.

In these expressions, the radial dependence on r has been separated from the angular (θ,ϕ)-dependence, which is given by the integral factors. Since these factors, play an important role in determining the *directional* properties of the radiated fields, we will denote them by the special notation:

$$\begin{aligned} Q(\boldsymbol{k}) &= \int_V \rho(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' \\ \boldsymbol{F}(\boldsymbol{k}) &= \int_V \boldsymbol{J}(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' \qquad \text{(radiation vector)} \end{aligned} \tag{14.7.5}$$

The first is also called the *charge form-factor*, and the second, the *radiation vector*. They are recognized to be the 3-dimensional spatial *Fourier transforms* of the charge and current densities. These quantities depend on ω or k and the directional unit vector $\hat{\mathbf{r}}$ which is completely defined by the spherical coordinate angles θ,ϕ. Therefore, whenever appropriate, we will indicate only the angular dependence in these quantities by writing them as $Q(\theta,\phi)$, $\boldsymbol{F}(\theta,\phi)$. In terms of this new notation, the far-field radiation potentials are:

$$\boxed{\begin{aligned} \varphi(\mathbf{r}) &= \frac{e^{-jkr}}{4\pi\epsilon r}\,Q(\theta,\phi) \\ \boldsymbol{A}(\mathbf{r}) &= \frac{\mu e^{-jkr}}{4\pi r}\,\boldsymbol{F}(\theta,\phi) \end{aligned}} \qquad \text{(radiation potentials)} \tag{14.7.6}$$

14.8 Radial Coordinates

The far-field solutions of Maxwell's equations and the directional patterns of antenna systems are best described in spherical coordinates.

The definitions of cartesian, cylindrical, and spherical coordinate systems are reviewed in Fig. 14.8.1 and are discussed further in Appendix E. The coordinates representing the vector $\mathbf{r}$ are, respectively, (x,y,z), (ρ,ϕ,z), and (r,θ,ϕ) and define orthogonal unit vectors in the corresponding directions, as shown in the figure.

The relationships between coordinate systems can be obtained by viewing the xy-plane and $z\rho$-plane, as shown in Fig. 14.8.2. The relationships between cartesian and cylindrical coordinates are:

$$\boxed{\begin{aligned} x &= \rho\cos\phi \\ y &= \rho\sin\phi \end{aligned}} \qquad \boxed{\begin{aligned} \hat{\boldsymbol{\rho}} &= \hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi \\ \hat{\boldsymbol{\phi}} &= -\hat{\mathbf{x}}\sin\phi + \hat{\mathbf{y}}\cos\phi \end{aligned}} \tag{14.8.1}$$

Similarly, the relationships of cylindrical to spherical coordinates are:

$$\boxed{\begin{aligned} \rho &= r\sin\theta \\ z &= r\cos\theta \end{aligned}} \qquad \boxed{\begin{aligned} \hat{\mathbf{r}} &= \hat{\mathbf{z}}\cos\theta + \hat{\boldsymbol{\rho}}\sin\theta \\ \hat{\boldsymbol{\theta}} &= -\hat{\mathbf{z}}\sin\theta + \hat{\boldsymbol{\rho}}\cos\theta \end{aligned}} \qquad \boxed{\begin{aligned} \hat{\mathbf{z}} &= \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta \\ \hat{\boldsymbol{\rho}} &= \hat{\mathbf{r}}\sin\theta + \hat{\boldsymbol{\theta}}\cos\theta \end{aligned}} \tag{14.8.2}$$

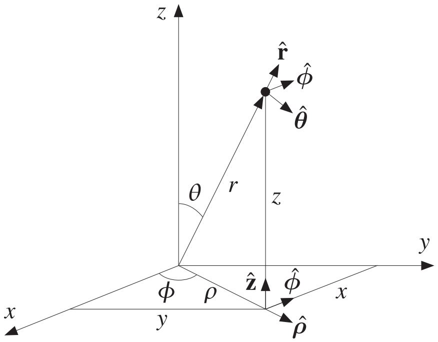


Fig. 14.8.1 Cartesian, cylindrical, and spherical coordinates.

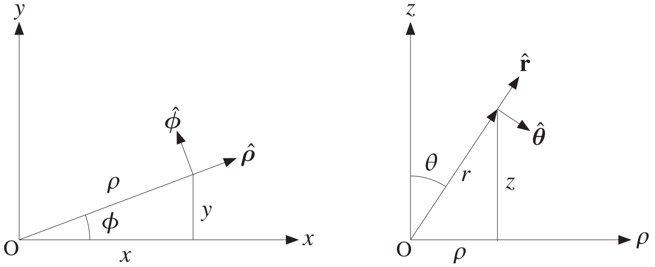


Fig. 14.8.2 Spherical coordinates viewed from xy-plane and $z\rho$-plane.

The relationships between cartesian and spherical coordinates are obtained from (14.8.2) by replacing ρ and $\hat{\boldsymbol{\rho}}$ in terms of Eq. (14.8.1), for example,

$$x = \rho\cos\phi = (r\sin\theta)\cos\phi = r\sin\theta\cos\phi$$

$$\hat{\mathbf{r}} = \hat{\boldsymbol{\rho}}\sin\theta + \hat{\mathbf{z}}\cos\theta = (\hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi)\sin\theta + \hat{\mathbf{z}}\cos\theta$$

The resulting relationships are:

$$\boxed{\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned}} \qquad \boxed{\begin{aligned} \hat{\mathbf{r}} &= \hat{\mathbf{x}}\cos\phi\sin\theta + \hat{\mathbf{y}}\sin\phi\sin\theta + \hat{\mathbf{z}}\cos\theta \\ \hat{\boldsymbol{\theta}} &= \hat{\mathbf{x}}\cos\phi\cos\theta + \hat{\mathbf{y}}\sin\phi\cos\theta - \hat{\mathbf{z}}\sin\theta \\ \hat{\boldsymbol{\phi}} &= -\hat{\mathbf{x}}\sin\phi + \hat{\mathbf{y}}\cos\phi \end{aligned}} \tag{14.8.3}$$

Note again that the radial unit vector $\hat{\mathbf{r}}$ is completely determined by the polar and azimuthal angles θ,ϕ. Infinitesimal length increments in each of the spherical unit-vector directions are defined by:

$$dl_r = dr\,,\quad dl_\theta = r\,d\theta\,,\quad dl_\phi = r\sin\theta\,d\phi \qquad \text{(spherical lengths)} \tag{14.8.4}$$

The gradient operator ∇ in spherical coordinates is:

$$\nabla = \hat{\mathbf{r}}\frac{\partial}{\partial l_r} + \hat{\boldsymbol{\theta}}\frac{\partial}{\partial l_\theta} + \hat{\boldsymbol{\phi}}\frac{\partial}{\partial l_\phi} = \hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial\theta} + \hat{\boldsymbol{\phi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi} \tag{14.8.5}$$

The lengths dl_θ and dl_ϕ correspond to infinitesimal displacements in the $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ directions on the surface of a sphere of radius r, as shown in Fig. 14.8.3. The surface element $d\mathbf{S} = \hat{\mathbf{r}}\,dS$ on the sphere is defined by $dS = dl_\theta\,dl_\phi$, or,

$$dS = r^2\sin\theta\,d\theta\,d\phi \tag{14.8.6}$$

The corresponding infinitesimal solid angle $d\Omega$ subtended by the $d\theta, d\phi$ cone is:

$$\boxed{dS = r^2 d\Omega} \quad\Rightarrow\quad \boxed{d\Omega = \frac{dS}{r^2} = \sin\theta\,d\theta\,d\phi} \tag{14.8.7}$$

The solid angle subtended by the whole sphere is in units of *steradians*:

$$\Omega_{\text{sphere}} = \int_0^\pi \sin\theta\,d\theta\int_0^{2\pi} d\phi = 4\pi$$

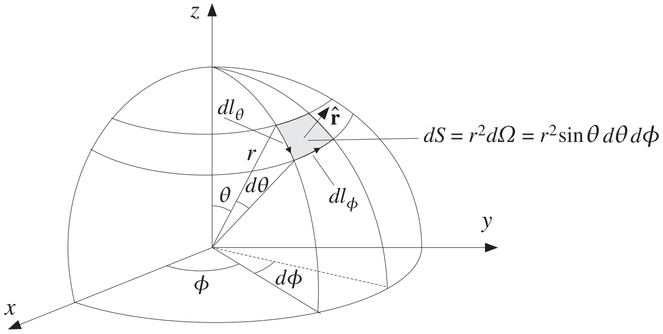


Fig. 14.8.3 Solid angle defined by angles θ,ϕ.

14.9 Radiation Field Approximation

In deriving the field intensities $\boldsymbol{E}$ and $\boldsymbol{H}$ from the far-field potentials (14.7.6), we must make one final approximation and keep only the terms that depend on r like $1/r$, and ignore terms that fall off faster, e.g., like $1/r^2$. We will refer to fields with $1/r$ dependence as *radiation fields*.

The justification for this approximation is shown in Fig. 14.9.1. The power radiated into a solid angle $d\Omega$ will flow through the surface area dS and will be given by $dP = \mathcal{P}_r dS$, where $\mathcal{P}_r$ is the radial component of the Poynting vector. Replacing dS in terms of the solid angle and $\mathcal{P}_r$ in terms of the squared electric field, we have:

$$dP = \mathcal{P}_r dS = \left(\frac{1}{2\eta}|E|^2\right)(r^2 d\Omega)$$

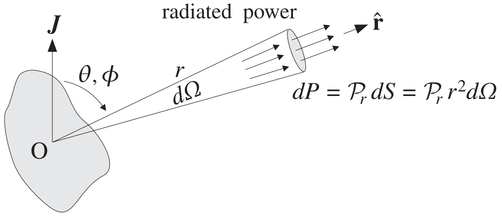


Fig. 14.9.1 Power radiated into solid angle $d\Omega$.

Thus, if the amount of power in the solid angle $d\Omega$ is to propagate away without attenuation with distance r, then the electric field must be such that $|E|^2 r^2 \sim$ const, or that $|E| \sim 1/r$; similarly, $|H| \sim 1/r$. Any terms in E, H that fall off faster than $1/r$ will not be capable of radiating power to large distances from their current sources.

14.10 Computing the Radiation Fields

At far distances from the localized current $\mathbf{J}$, the radiation fields can be obtained from Eqs. (14.3.9) by using the radiation vector potential $\mathbf{A}$ of Eq. (14.7.6). In computing the curl of $\mathbf{A}$, we may ignore any terms that fall off faster than $1/r$:

$$\nabla\times\mathbf{A} = \nabla\times\left(\frac{\mu e^{-jkr}}{4\pi r}\mathbf{F}\right) = \left(\hat{\mathbf{r}}\,\frac{\partial}{\partial r} + \text{angular derivatives}\right)\times\left(\frac{\mu e^{-jkr}}{4\pi r}\mathbf{F}\right)$$

$$= -jk\,(\hat{\mathbf{r}}\times\mathbf{F})\left(\frac{\mu e^{-jkr}}{4\pi r}\right) + O\left(\frac{1}{r^2}\right) = -j\,\mathbf{k}\times\mathbf{A} + O\left(\frac{1}{r^2}\right)$$

The "angular derivatives" arise from the θ,ϕ derivatives in the gradient as per Eq. (14.8.5). These derivatives act on $\mathbf{F}(\theta,\phi)$, but because they already have a $1/r$ factor in them and the rest of $\mathbf{A}$ has another $1/r$ factor, these terms will go down like $1/r^2$. Similarly, when we compute the derivative $\partial_r[e^{-jkr}/r]$ we may keep only the derivative of the numerator because the rest goes down like $1/r^2$.

Thus, we arrive at the useful *rule* that to order $1/r$, the gradient operator ∇, *whenever* it acts on a function of the form $f(\theta,\phi)e^{-jkr}/r$, can be replaced by:

$$\boxed{\nabla \longrightarrow -j\,\mathbf{k} = -jk\,\hat{\mathbf{r}}} \tag{14.10.1}$$

Applying the rule again, we have:

$$\nabla\times(\nabla\times\mathbf{A}) = -j\,\mathbf{k}\times(-j\,\mathbf{k}\times\mathbf{A}) = (\mathbf{k}\times\mathbf{A})\times\mathbf{k} = k^2(\hat{\mathbf{r}}\times\mathbf{A})\times\hat{\mathbf{r}} = \omega^2\mu\epsilon\,(\hat{\mathbf{r}}\times\mathbf{A})\times\hat{\mathbf{r}}$$

Noting that $\omega\mu = ck\mu = k\sqrt{\mu/\epsilon} = k\eta$ and using Eq. (14.3.9), we finally find:

$$\boxed{\begin{aligned} \mathbf{E} &= -jk\eta\,\frac{e^{-jkr}}{4\pi r}\,(\hat{\mathbf{r}}\times\mathbf{F})\times\hat{\mathbf{r}} \\ \mathbf{H} &= -jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\mathbf{F} \end{aligned}} \qquad \text{(radiation fields)} \tag{14.10.2}$$

Moreover, we recognize that:

$$\boxed{\mathbf{E} = \eta\mathbf{H}\times\hat{\mathbf{r}}, \quad \mathbf{H} = \frac{1}{\eta}\hat{\mathbf{r}}\times\mathbf{E}} \qquad \text{and} \qquad \frac{|\mathbf{E}|}{|\mathbf{H}|} = \eta \tag{14.10.3}$$

We note the similarity to uniform plane waves and emphasize the following properties:

1. $\{\mathbf{E},\, \mathbf{H},\, \hat{\mathbf{r}}\}$ form a right-handed vector system.
2. $\mathbf{E}$ is always parallel to the transverse part $\mathbf{F}_\perp$ of the radiation vector $\mathbf{F}$.
3. $\mathbf{H}$ is always perpendicular to the radiation vector $\mathbf{F}$.
4. dc current sources $(\omega = k = 0)$ will not radiate.

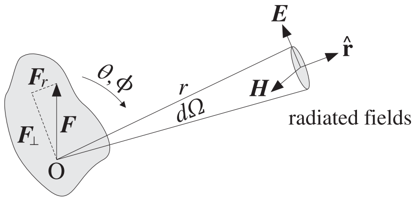


Fig. 14.10.1 Electric and magnetic fields radiated by a current source.

Figure 14.10.1 illustrates some of these remarks. The radiation vector may be decomposed in general into a radial part $\mathbf{F}_r = \hat{\mathbf{r}}F_r$ and a transverse part $\mathbf{F}_\perp$. In fact, this decomposition is obtained from the identity:

$$\mathbf{F} = \hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{F}) + (\hat{\mathbf{r}}\times\mathbf{F})\times\hat{\mathbf{r}} = \hat{\mathbf{r}}F_r + \mathbf{F}_\perp$$

Resolving $\mathbf{F}$ along the spherical coordinate unit vectors, we have:

$$\begin{aligned} \mathbf{F} &= \hat{\mathbf{r}}F_r + \hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi \\ \hat{\mathbf{r}}\times\mathbf{F} &= \hat{\boldsymbol{\phi}}F_\theta - \hat{\boldsymbol{\theta}}F_\phi \\ \mathbf{F}_\perp &= (\hat{\mathbf{r}}\times\mathbf{F})\times\hat{\mathbf{r}} = \hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi \end{aligned}$$

Thus, only F_θ and F_ϕ contribute to the fields:

$$\boxed{\begin{aligned} \mathbf{E} &= -jk\eta\,\frac{e^{-jkr}}{4\pi r}\,\big[\hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi\big] \\ \mathbf{H} &= -jk\,\frac{e^{-jkr}}{4\pi r}\,\big[\hat{\boldsymbol{\phi}}F_\theta - \hat{\boldsymbol{\theta}}F_\phi\big] \end{aligned}} \qquad \text{(radiation fields)} \tag{14.10.4}$$

Recognizing that $\hat{\mathbf{r}}\times\mathbf{F} = \hat{\mathbf{r}}\times\mathbf{F}_\perp$, we can also write compactly:

$$\boxed{\begin{aligned} \mathbf{E} &= -jk\eta\,\frac{e^{-jkr}}{4\pi r}\,\mathbf{F}_\perp \\ \mathbf{H} &= -jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\mathbf{F}_\perp \end{aligned}} \qquad \text{(radiation fields)} \tag{14.10.5}$$

In general, the radiation vector will have both F_θ and F_ϕ components, depending on the nature of the current distribution $\boldsymbol{J}$. However, in practice there are three important cases that stand out:

1. Only F_θ is present. This includes all *linear antennas* and arrays. The z-axis is oriented in the direction of the antenna, so that the radiation vector only has r and θ components.
2. Only F_ϕ is present. This includes *loop antennas* with the xy-plane chosen as the plane of the loop.
3. Both F_θ and F_ϕ are present, but they are carefully chosen to have the phase relationship $F_\phi = \pm jF_\theta$, so that the resulting electric field will be *circularly polarized*. This includes *helical antennas* used in space communications.

14.11 Problems

14.1 First, prove the differential identity:

$$\nabla' \cdot \big[\boldsymbol{J}(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}\big] = j\boldsymbol{k}\cdot\boldsymbol{J}(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'} - j\omega\rho(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}$$

Then, prove the integral identity:

$$\boldsymbol{k}\cdot\int_V \boldsymbol{J}(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' = \omega\int_V \rho(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}'$$

Assume that the charge and current densities are localized within the finite volume V. Finally, show that the charge form-factor Q and radiation vector $\boldsymbol{F}$ are related by:

$$\hat{\mathbf{r}}\cdot\boldsymbol{F} = cQ$$

14.2 Using similar techniques as in the previous problem, prove the following general property, valid for any scalar function $g(\mathbf{r})$, where V is the volume over which $\boldsymbol{J},\rho$ are non-zero:

$$\int_V \boldsymbol{J}(\mathbf{r}')\cdot\nabla' g(\mathbf{r}')\,d^3\mathbf{r}' = j\omega\int_V g(\mathbf{r}')\rho(\mathbf{r}')\,d^3\mathbf{r}'$$

14.3 It is possible to obtain the fields generated by the source densities $\rho,\boldsymbol{J}$ by working directly with Maxwell's equations without introducing the scalar and vector potentials $\phi,\boldsymbol{A}$. Start with the monochromatic Maxwell's equations

$$\nabla\times\boldsymbol{E} = -j\omega\mu\boldsymbol{H},\quad \nabla\times\boldsymbol{H} = \boldsymbol{J} + j\omega\epsilon\boldsymbol{E},\quad \nabla\cdot\boldsymbol{E} = \frac{1}{\epsilon}\rho,\quad \nabla\cdot\boldsymbol{H} = 0$$

Show that $\boldsymbol{E},\boldsymbol{H}$ satisfy the following Helmholtz equations:

$$(\nabla^2 + k^2)\boldsymbol{E} = j\omega\mu\boldsymbol{J} + \frac{1}{\epsilon}\nabla\rho,\quad (\nabla^2 + k^2)\boldsymbol{H} = -\nabla\times\boldsymbol{J}$$

Show that their solutions are obtained with the help of the Green's function (14.3.4):

$$\boldsymbol{E} = \int_V \big[-j\omega\mu\boldsymbol{J}G - \frac{1}{\epsilon}(\nabla'\rho)G\big]\,dV'$$

$$\boldsymbol{H} = \int_V \big[\nabla'\times\boldsymbol{J}\big]G\,dV'$$

Although these expressions and Eqs. (14.3.10) look slightly different, they are equivalent. Explain in what sense this is true.

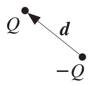
14.4 The electric and magnetic dipole moments of charge and current volume distributions $\rho,\boldsymbol{J}$ are defined by:

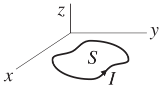
$$\mathbf{p} = \int_V \mathbf{r}\,\rho(\mathbf{r})\,dV\,,\quad \mathbf{m} = \frac{1}{2}\int_V \mathbf{r}\times\boldsymbol{J}(\mathbf{r})\,dV$$

Using these definitions and the integral property of Eq. (C.41) of Appendix C, show that for two charges $\pm Q$ separated by distance $\mathbf{d}$, and for a current I flowing on a closed planar loop of arbitrary shape and area S lying on the xy-plane, the quantities $\mathbf{p},\mathbf{m}$ are given by:

$$\mathbf{p} = Q\,\mathbf{d}$$

$$\mathbf{m} = \hat{\mathbf{z}}\,I\,S$$





14.5 By performing an inverse Fourier time transform on Eq. (14.5.5), show that the fields produced by an arbitrary time-varying dipole at the origin, $\boldsymbol{P}(\mathbf{r},t) = \mathbf{p}(t)\,\delta^{(3)}(\mathbf{r})$, are given by:

$$\boldsymbol{E}(\mathbf{r},t) = \frac{1}{\epsilon_0}\Big(\frac{1}{c_0}\frac{\partial}{\partial t} + \frac{1}{r}\Big)\left[\frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{p}(t_r)) - \mathbf{p}(t_r)}{r}\right]\frac{1}{4\pi r} - \frac{1}{\epsilon_0 c_0^2}\,\hat{\mathbf{r}}\times(\ddot{\mathbf{p}}(t_r)\times\hat{\mathbf{r}})\,\frac{1}{4\pi r}$$

$$\boldsymbol{H}(\mathbf{r},t) = \frac{\partial}{\partial t}\Big(\frac{1}{c_0}\frac{\partial}{\partial t} + \frac{1}{r}\Big)(\mathbf{p}(t_r)\times\hat{\mathbf{r}})\,\frac{1}{4\pi r}$$

where $t_r = t - r/c_0$ is the retarded time and the time-derivatives act only on $\mathbf{p}(t_r)$. Show also that the radiated fields are (with $\eta_0 = \sqrt{\mu_0/\epsilon_0}$):

$$\boldsymbol{E}_{\text{rad}}(\mathbf{r},t) = \mu_0\,\hat{\mathbf{r}}\times(\hat{\mathbf{r}}\times\ddot{\mathbf{p}}(t_r))\,\frac{1}{4\pi r} = \eta_0\,\boldsymbol{H}_{\text{rad}}(\mathbf{r},t)\times\hat{\mathbf{r}}$$

$$\boldsymbol{H}_{\text{rad}}(\mathbf{r},t) = \frac{\mu_0}{\eta_0}(\ddot{\mathbf{p}}(t_r)\times\hat{\mathbf{r}})\,\frac{1}{4\pi r}$$

14.6 Assume that the dipole of the previous problem is along the z-direction, $\mathbf{p}(t) = \hat{\mathbf{z}}\,p(t)$. Integrating the Poynting vector $\boldsymbol{\mathcal{P}} = \boldsymbol{E}_{\text{rad}}\times\boldsymbol{H}_{\text{rad}}$ over a sphere of radius r, show that the total radiated power from the dipole is given by:

$$P_{\text{rad}}(r,t) = \frac{\eta_0}{6\pi c_0^2}\,\ddot{p}^2(t_r)$$

14.7 Define a 3×3 matrix $J(\mathbf{a})$ such that the operation $J(\mathbf{a})\mathbf{b}$ represents the cross-product $\mathbf{a}\times\mathbf{b}$. Show that:

$$J(\mathbf{a}) = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}$$

Show that $J(\mathbf{a})$ is a rank-2 matrix with eigenvalues $\lambda = 0$ and $\lambda = \pm j|\mathbf{a}|$, where $\mathbf{a}$ is assumed to be real-valued. Show that the eigenvectors corresponding to the non-zero eigenvalues are given by $\mathbf{e} = \hat{\mathbf{f}} \mp j\,\hat{\mathbf{g}}$, where $\hat{\mathbf{f}},\hat{\mathbf{g}}$ are real-valued unit vectors such that $\{\hat{\mathbf{f}},\hat{\mathbf{g}},\hat{\mathbf{a}}\}$ is a right-handed vector system (like $\{\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}\}$), here, $\hat{\mathbf{a}} = \mathbf{a}/|\mathbf{a}|$. Show that $\mathbf{e}\cdot\mathbf{e} = 0$ and $\mathbf{e}^*\cdot\mathbf{e} = 2$.

A radiator consists of electric and magnetic dipoles $\mathbf{p},\mathbf{m}$ placed at the origin. Assuming harmonic time dependence and adding the radiation fields of Eqs. (14.5.6) and (14.5.10), show that the total radiated fields can be expressed in terms of the 6×6 matrix operation:

$$\begin{bmatrix} \boldsymbol{E}(\mathbf{r}) \\ \eta_0\boldsymbol{H}(\mathbf{r}) \end{bmatrix} = -\eta_0\,k^2\,\frac{e^{-jkr}}{4\pi r}\begin{bmatrix} J^2(\hat{\mathbf{r}}) & J(\hat{\mathbf{r}}) \\ -J(\hat{\mathbf{r}}) & J^2(\hat{\mathbf{r}}) \end{bmatrix}\begin{bmatrix} c_0\,\mathbf{p} \\ \mathbf{m} \end{bmatrix}$$

Show that $J(\hat{\mathbf{r}})$ satisfies the matrix equation $J^3(\hat{\mathbf{r}})+J(\hat{\mathbf{r}})=0$. Moreover, show that its eigenvalues are $\lambda = 0$ and $\lambda = \pm j$ and that the eigenvectors belonging to the two nonzero eigenvalues are given in terms of the polar unit vectors by $\mathbf{e} = \hat{\boldsymbol{\theta}} \mp j\,\hat{\boldsymbol{\phi}}$.

Because the matrix $J(\hat{\mathbf{r}})$ is rank-defective, so is the above 6×6 matrix, reflecting the fact that the radiation fields can only have two polarization states. However, it has been shown recently [1003] that in a multiple-scattering environment, such as wireless propagation in cities, the corresponding 6×6 matrix becomes a full-rank matrix (rank 6) allowing the tripling of the channel capacity over the standard dual-polarization transmission.

15

Transmitting and Receiving Antennas

15.1 Energy Flux and Radiation Intensity

The flux of electromagnetic energy radiated from a current source at far distances is given by the time-averaged Poynting vector, calculated in terms of the radiation fields (14.10.4):

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\mathrm{Re}(\boldsymbol{E}\times\boldsymbol{H}^*) = \frac{1}{2}\left(-jk\eta\,\frac{e^{-jkr}}{4\pi r}\right)\left(jk\,\frac{e^{jkr}}{4\pi r}\right)\mathrm{Re}\big[\,(\hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi)\times(\hat{\boldsymbol{\phi}}F_\theta^* - \hat{\boldsymbol{\theta}}F_\phi^*)\,\big]$$

Noting that $\hat{\boldsymbol{\theta}}\times\hat{\boldsymbol{\phi}} = \hat{\mathbf{r}}$, we have:

$$(\hat{\boldsymbol{\theta}}F_\theta + \hat{\boldsymbol{\phi}}F_\phi)\times(\hat{\boldsymbol{\phi}}F_\theta^* - \hat{\boldsymbol{\theta}}F_\phi^*) = \hat{\mathbf{r}}\big(|F_\theta|^2 + |F_\phi|^2\big) = \hat{\mathbf{r}}\,\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2$$

Therefore, the energy flux vector will be:

$$\boldsymbol{\mathcal{P}} = \hat{\mathbf{r}}\,\mathcal{P}_r = \hat{\mathbf{r}}\,\frac{\eta k^2}{32\pi^2 r^2}\,\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2 \tag{15.1.1}$$

Thus, the radiated energy flows radially away from the current source and attenuates with the square of the distance. The *angular distribution* of the radiated energy is described by the *radiation pattern* factor:

$$\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2 = \big|F_\theta(\theta,\phi)\big|^2 + \big|F_\phi(\theta,\phi)\big|^2 \tag{15.1.2}$$

With reference to Fig. 14.9.1, the power dP intercepting the area element $dS = r^2 d\Omega$ defines the power per unit area, or the *power density* of the radiation:

$$\boxed{\frac{dP}{dS} = \frac{dP}{r^2 d\Omega} = \mathcal{P}_r = \frac{\eta k^2}{32\pi^2 r^2}\,\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2} \qquad \text{(power density)} \tag{15.1.3}$$

The *radiation intensity* $U(\theta,\phi)$ is defined to be the power radiated per unit solid angle, that is, the quantity $dP/d\Omega = r^2 dP/dS = r^2\mathcal{P}_r$:

$$\boxed{U(\theta,\phi) = \frac{dP}{d\Omega} = r^2\mathcal{P}_r = \frac{\eta k^2}{32\pi^2}\,\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2} \qquad \text{(radiation intensity)} \tag{15.1.4}$$

The total *radiated power* is obtained by integrating Eq. (15.1.4) over all solid angles $d\Omega = \sin\theta\, d\theta d\phi$, that is, over $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$:

$$\boxed{P_{\text{rad}} = \int_0^{\pi}\int_0^{2\pi} U(\theta,\phi)\, d\Omega} \qquad \text{(total radiated power)} \tag{15.1.5}$$

A useful concept is that of an *isotropic radiator*—a radiator whose intensity is the same in all directions. In this case, the total radiated power P_{rad} will be equally distributed over all solid angles, that is, over the total solid angle of a sphere $\Omega_{\text{sphere}} = 4\pi$ steradians, and therefore, the *isotropic radiation intensity* will be:

$$U_I = \left(\frac{dP}{d\Omega}\right)_I = \frac{P_{\text{rad}}}{\Omega_{\text{sphere}}} = \frac{P_{\text{rad}}}{4\pi} = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} U(\theta,\phi)\, d\Omega \tag{15.1.6}$$

Thus, U_I is the *average* of the radiation intensity over all solid angles. The corresponding power density of such an isotropic radiator will be:

$$\left(\frac{dP}{dS}\right)_I = \frac{U_I}{r^2} = \frac{P_{\text{rad}}}{4\pi r^2} \qquad \text{(isotropic power density)} \tag{15.1.7}$$

15.2 Directivity, Gain, and Beamwidth

The *directive gain* of an antenna system towards a given direction (θ,ϕ) is the radiation intensity normalized by the corresponding *isotropic* intensity, that is,

$$\boxed{D(\theta,\phi) = \frac{U(\theta,\phi)}{U_I} = \frac{U(\theta,\phi)}{P_{\text{rad}}/4\pi} = \frac{4\pi}{P_{\text{rad}}}\frac{dP}{d\Omega}} \qquad \text{(directive gain)} \tag{15.2.1}$$

It measures the ability of the antenna to direct its power towards a given direction. The maximum value of the directive gain, $D_{\max}$, is called the *directivity* of the antenna and will be realized towards some particular direction, say (θ_0,ϕ_0). The radiation intensity will be maximum towards that direction, $U_{\max} = U(\theta_0,\phi_0)$, so that

$$\boxed{D_{\max} = \frac{U_{\max}}{U_I}} \qquad \text{(directivity)} \tag{15.2.2}$$

The directivity is often expressed in dB,[†] that is, $D_{\text{dB}} = 10\log_{10} D_{\max}$. Re-expressing the radiation intensity in terms of the directive gain, we have:

$$\frac{dP}{d\Omega} = U(\theta,\phi) = D(\theta,\phi)\,U_I = \frac{P_{\text{rad}} D(\theta,\phi)}{4\pi} \tag{15.2.3}$$

and for the power density in the direction of (θ,ϕ):

$$\frac{dP}{dS} = \frac{dP}{r^2 d\Omega} = \frac{P_{\text{rad}} D(\theta,\phi)}{4\pi r^2} \qquad \text{(power density)} \tag{15.2.4}$$

[†] The term “dBi” is often used as a reminder that the directivity is with respect to the isotropic case.

Comparing with Eq. (15.1.7), we note that if the amount of power $P_{\text{rad}} D(\theta,\phi)$ were emitted isotropically, then Eq. (15.2.4) would be the corresponding isotropic power density. Therefore, we will refer to $P_{\text{rad}} D(\theta,\phi)$ as the *effective isotropic power*, or the *effective radiated power* (ERP) towards the (θ,ϕ)-direction.

In the direction of maximum gain, the quantity $P_{\text{rad}} D_{\max}$ will be referred to as the *effective isotropic radiated power* (EIRP). It defines the maximum power density achieved by the antenna:

$$\boxed{\left(\frac{dP}{dS}\right)_{\max} = \frac{P_{\text{EIRP}}}{4\pi r^2}}, \qquad \text{where} \quad P_{\text{EIRP}} = P_{\text{rad}} D_{\max} \tag{15.2.5}$$

Usually, communicating antennas—especially highly directive ones such as dish antennas—are oriented to point towards the maximum directive gain of each other.

A related concept is that of the *power gain*, or simply the *gain* of an antenna. It is defined as in Eq. (15.2.1), but instead of being normalized by the total radiated power, it is normalized to the total power P_T *accepted by the antenna terminals* from a connected transmitter, as shown in Fig. 15.2.1:

$$\boxed{G(\theta,\phi) = \frac{U(\theta,\phi)}{P_T/4\pi} = \frac{4\pi}{P_T}\frac{dP}{d\Omega}} \qquad \text{(power gain)} \tag{15.2.6}$$

We will see in Sec. 15.4 that the power P_T delivered to the antenna terminals is at most half the power produced by the generator—the other half being dissipated as heat in the generator’s internal resistance.

Moreover, the power P_T may differ from the power radiated, P_{rad}, because of several loss mechanisms, such as ohmic losses of the currents flowing on the antenna wires or losses in the dielectric surrounding the antenna.

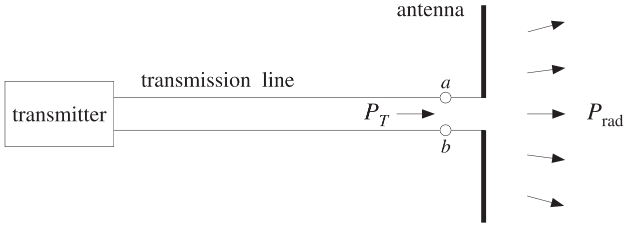


Fig. 15.2.1 Power delivered to an antenna versus power radiated.

The definition of power gain does not include any reflection losses arising from improper matching of the transmission line to the antenna input impedance [92]. The *efficiency factor* of the antenna is defined by:

$$e = \frac{P_{\text{rad}}}{P_T} \qquad \Rightarrow \qquad P_{\text{rad}} = eP_T \tag{15.2.7}$$

In general, $0 \le e \le 1$. For a lossless antenna the efficiency factor will be unity and $P_{\text{rad}} = P_T$. In such an ideal case, there is no distinction between directive and power gain. Using Eq. (15.2.7) in (15.2.1), we find $G = 4\pi U/P_T = e4\pi U/P_{\text{rad}}$, or,

$$\boxed{G(\theta,\phi)=eD(\theta,\phi)} \tag{15.2.8}$$

The *maximum gain* is related to the directivity by $G_{\max} = eD_{\max}$. It follows that the effective radiated power can be written as $P_{\text{rad}}D(\theta,\phi)=P_T G(\theta,\phi)$, and the EIRP, $P_{\text{EIRP}} = P_{\text{rad}}D_{\max} = P_T G_{\max}$.

The angular distribution functions we defined thus far, that is, $G(\theta,\phi)$, $D(\theta,\phi)$, $U(\theta,\phi)$ are all proportional to each other. Each brings out a different aspect of the radiating system. In describing the angular distribution of radiation, it proves convenient to consider it relative to its maximal value. Thus, we define the normalized power pattern, or *normalized gain* by:

$$\boxed{g(\theta,\phi)=\frac{G(\theta,\phi)}{G_{\max}}} \qquad \text{(normalized gain)} \tag{15.2.9}$$

Because of the proportionality of the various angular functions, we have:

$$g(\theta,\phi)=\frac{G(\theta,\phi)}{G_{\max}}=\frac{D(\theta,\phi)}{D_{\max}}=\frac{U(\theta,\phi)}{U_{\max}}=\frac{\big|\boldsymbol{F}_\perp(\theta,\phi)\big|^2}{|\boldsymbol{F}_\perp|^2_{\max}} \tag{15.2.10}$$

Writing $P_T G(\theta,\phi)=P_T G_{\max}\, g(\theta,\phi)$, we have for the power density:

$$\frac{dP}{dS}=\frac{P_T G_{\max}}{4\pi r^2}\,g(\theta,\phi)=\frac{P_{\text{EIRP}}}{4\pi r^2}\,g(\theta,\phi) \tag{15.2.11}$$

This form is useful for describing communicating antennas and radar. The normalized gain is usually displayed in a polar plot with polar coordinates (ρ,θ) such that $\rho = g(\theta)$, as shown in Fig. 15.2.2. (This figure depicts the gain of a half-wave dipole antenna given by $g(\theta)=\cos^2(0.5\pi\cos\theta)/\sin^2\theta$.) The 3-dB, or half-power, beamwidth is defined as the difference $\Delta\theta_B = \theta_2-\theta_1$ of the 3-dB angles at which the normalized gain is equal to 1/2, or, −3 dB.

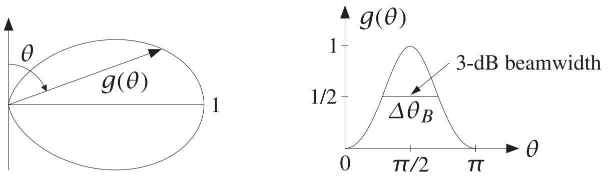


Fig. 15.2.2 Polar and regular plots of normalized gain versus angle.

The MATLAB functions `dbp`, `abp`, `dbz`, `abz` given in Appendix I allow the plotting of the gain in dB or in absolute units versus the polar angle θ or the azimuthal angle ϕ. Their typical usage is as follows:

```
dbp(theta, g, rays, Rm, width);     % polar gain plot in dB
abp(theta, g, rays, width);         % polar gain plot in absolute units
dbz(phi, g, rays, Rm, width);       % azimuthal gain plot in dB
abz(phi, g, rays, width);           % azimuthal gain plot in absolute units
```

Example 15.2.1: A TV station is transmitting 10 kW of power with a gain of 15 dB towards a particular direction. Determine the peak and rms value of the electric field E at a distance of 5 km from the station.

Solution: The gain in absolute units will be $G = 10^{G_{\text{dB}}/10} = 10^{15/10} = 31.62$. It follows that the radiated EIRP will be $P_{\text{EIRP}} = P_T G = 10\times 31.62 = 316.2$ kW. The electric field at distance $r = 5$ km is obtained from Eq. (15.2.5):

$$\frac{dP}{dS}=\frac{P_{\text{EIRP}}}{4\pi r^2}=\frac{1}{2\eta}E^2 \quad\Rightarrow\quad E=\frac{1}{r}\sqrt{\frac{\eta P_{\text{EIRP}}}{2\pi}}$$

This gives $E = 0.87$ V/m. The rms value is $E_{\text{rms}} = E/\sqrt{2} = 0.62$ V/m. □

Another useful concept is that of the *beam solid angle* of an antenna. The definition is motivated by the case of a highly directive antenna, which concentrates all of its radiated power P_{rad} into a small solid angle $\Delta\Omega$, as illustrated in Fig. 15.2.3.

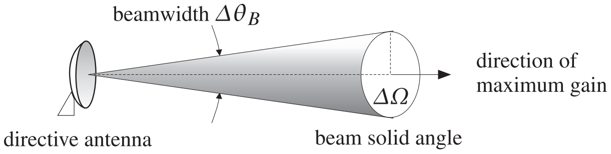


Fig. 15.2.3 Beam solid angle and beamwidth of a highly directive antenna.

The radiation intensity in the direction of the solid angle will be:

$$U=\frac{\Delta P}{\Delta\Omega}=\frac{P_{\text{rad}}}{\Delta\Omega} \tag{15.2.12}$$

where $\Delta P = P_{\text{rad}}$ by assumption. It follows that: $D_{\max} = 4\pi U/P_{\text{rad}} = 4\pi/\Delta\Omega$, or,

$$\boxed{D_{\max}=\frac{4\pi}{\Delta\Omega}} \tag{15.2.13}$$

Thus, the more concentrated the beam, the higher the directivity. Although (15.2.13) was derived under the assumption of a highly directive antenna, it may be used as the *definition* of the beam solid angle for any antenna, that is,

$$\boxed{\Delta\Omega=\frac{4\pi}{D_{\max}}} \qquad \text{(beam solid angle)} \tag{15.2.14}$$

Using $D_{\max} = U_{\max}/U_I$ and Eq. (15.1.6), we have

$$\Delta\Omega=\frac{4\pi U_I}{U_{\max}}=\frac{1}{U_{\max}}\int_0^{\pi}\!\!\int_0^{2\pi}U(\theta,\phi)\,d\Omega\,, \qquad \text{or,}$$

$$\boxed{\Delta\Omega=\int_0^{\pi}\!\!\int_0^{2\pi}g(\theta,\phi)\,d\Omega} \qquad \text{(beam solid angle)} \tag{15.2.15}$$

where $g(\theta,\phi)$ is the normalized gain of Eq. (15.2.10). Writing $P_{\text{rad}} = 4\pi U_I$, we have:

$$\Delta\Omega = \frac{P_{\text{rad}}}{U_{\text{max}}} \quad\Rightarrow\quad U_{\text{max}} = \frac{P_{\text{rad}}}{\Delta\Omega} \tag{15.2.16}$$

This is the general case of Eq. (15.2.12). We can also write:

$$P_{\text{rad}} = U_{\text{max}}\Delta\Omega \tag{15.2.17}$$

This is convenient for the numerical evaluation of P_{rad}. To get a measure of the beamwidth of a highly directive antenna, we assume that the directive gain is equal to its maximum *uniformly* over the entire solid angle $\Delta\Omega$ in Fig. 15.2.3, that is, $D(\theta,\phi) = D_{\text{max}}$, for $0 \le \theta \le \Delta\theta_B/2$. This implies that the normalized gain will be:

$$g(\theta,\phi) = \begin{cases} 1, & \text{if} \quad 0 \le \theta \le \Delta\theta_B/2 \\ 0, & \text{if} \quad \Delta\theta_B/2 < \theta \le \pi \end{cases}$$

Then, it follows from the definition (15.2.15) that:

$$\Delta\Omega = \int_0^{\Delta\theta_B/2}\int_0^{2\pi} d\Omega = \int_0^{\Delta\theta_B/2}\int_0^{2\pi} \sin\theta\, d\theta\, d\phi = 2\pi\left(1 - \cos\frac{\Delta\theta_B}{2}\right) \tag{15.2.18}$$

Using the approximation $\cos x \simeq 1 - x^2/2$, we obtain for small beamwidths:

$$\boxed{\Delta\Omega = \frac{\pi}{4}(\Delta\theta_B)^2} \tag{15.2.19}$$

and therefore the directivity can be expressed in terms of the beamwidth:

$$\boxed{D_{\text{max}} = \frac{16}{\Delta\theta_B^2}} \tag{15.2.20}$$

Example 15.2.2: Find the beamwidth in degrees of a lossless dish antenna with gain of 15 dB. The directivity and gain are equal in this case, therefore, Eq. (15.2.20) can be used to calculate the beamwidth: $\Delta\theta_B = \sqrt{16/D}$, where $D = G = 10^{15/10} = 31.62$. We find $\Delta\theta_B = 0.71$ rads, or $\Delta\theta_B = 40.76^{\text{o}}$.

For an antenna with 40 dB gain/directivity, we would have $D = 10^4$ and find $\Delta\theta_B = 0.04$ rads $= 2.29^{\text{o}}$. ☐

Example 15.2.3: A satellite in a geosynchronous orbit of 36,000 km is required to have complete earth coverage. What is its antenna gain in dB and its beamwidth? Repeat if the satellite is required to have coverage of an area equal the size of continental US.

Solution: The radius of the earth is $R = 6400$ km. Looking down from the satellite the earth appears as a flat disk of area $\Delta S = \pi R^2$. It follows that the subtended solid angle and the corresponding directivity/gain will be:

$$\Delta\Omega = \frac{\Delta S}{r^2} = \frac{\pi R^2}{r^2} \quad\Rightarrow\quad D = \frac{4\pi}{\Delta\Omega} = \frac{4r^2}{R^2}$$

With $r = 36{,}000$ km and $R = 6400$ km, we find $D = 126.56$ and in dB, $D_{\text{dB}} = 10\log_{10}D = 21.02$ dB. The corresponding beamwidth will be $\Delta\theta_B = \sqrt{16/D} = 0.36$ rad $= 20.37^{\text{o}}$.

For the continental US, the coast-to-coast distance of 3000 mi, or 4800 km, translates to an area of radius $R = 2400$ km, which leads to $D = 900$ and $D_{\text{dB}} = 29.54$ dB. The beamwidth is in this case $\Delta\theta_B = 7.64^{\text{o}}$.

Viewing the earth as a flat disk overestimates the required angle $\Delta\theta_B$ for earth coverage. Looking down from a satellite at a height r, the angle between the vertical and the tangent to the earth's surface is given by $\sin\theta = R/(r+R)$, which gives for $r = 36{,}000$ km, $\theta = 8.68^{\text{o}}$. The subtended angle will be then $\Delta\theta_B = 2\theta = 0.303$ rad $= 17.36^{\text{o}}$. It follows that the required antenna gain should be $G = 16/\Delta\theta_B^2 = 174.22 = 22.41$ dB. The flat-disk approximation is more accurate for smaller areas on the earth's surface that lie directly under the satellite. ☐

Example 15.2.4: The radial distance of a geosynchronous orbit can be calculated by equating centripetal and gravitational accelerations, and requiring that the angular velocity of the satellite corresponds to the period of 1 day, that is, $\omega = 2\pi/T$, where $T = 24$ hr $= 86\,400$ sec. Let m be the mass of the satellite and $M_\oplus$ the mass of the earth (see Appendix A):

$$\frac{GmM_\oplus}{r^2} = m\omega^2 r = m\left(\frac{2\pi}{T}\right)^2 r \quad\Rightarrow\quad r = \left(\frac{GM_\oplus T^2}{4\pi^2}\right)^{1/3}$$

The distance r is measured from the Earth's center. The corresponding height from the surface of the Earth is $h = r - R$. For the more precise value of $R = 6378$ km, the calculated values are:

$r = 42\,237$ km $= 26\,399$ mi
$h = 35\,860$ km $= 22\,414$ mi

15.3 Effective Area

When an antenna is operating as a receiving antenna, it extracts a certain amount of power from an incident electromagnetic wave. As shown in Fig. 15.3.1, an incident wave coming from a far distance may be thought of as a uniform plane wave being intercepted by the antenna.

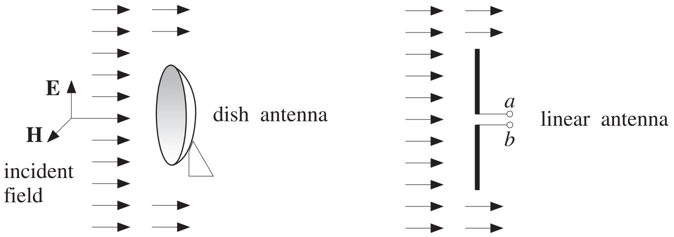


Fig. 15.3.1 Effective area of an antenna.

The incident electric field sets up currents on the antenna. Such currents may be represented by a Thévenin-equivalent generator, which delivers power to any connected receiving load impedance.

The induced currents also re-radiate an electric field (referred to as the scattered field), which interferes with the incident field causing a shadow region behind the antenna, as shown in Fig. 15.3.1.

The total electric field outside the antenna will be the sum of the incident and reradiated fields. For a perfectly conducting antenna, the boundary conditions are that the tangential part of the total electric field vanish on the antenna surface. In Chap. 21, we apply these boundary conditions to obtain and solve Hallén's and Pocklington's integral equations satisfied by the induced current.

The power density of the incident wave at the location of the receiving antenna can be expressed in terms of the electric field of the wave, $\mathcal{P}_{\text{inc}} = E^2/2\eta$.

The *effective area* or *effective aperture* A of the antenna is defined to be that area which when intercepted by the incident power density $\mathcal{P}_{\text{inc}}$ gives the amount of *received power* P_R available at the antenna output terminals [92]:

$$\boxed{P_R = A\mathcal{P}_{\text{inc}}} \tag{15.3.1}$$

For a lossy antenna, the available power at the terminals will be somewhat less than the extracted radiated power P_{rad}, by the efficiency factor $P_R = eP_{\text{rad}}$. Thus, we may also define the *maximum effective aperture* A_m as the area which extracts the power P_{rad} from the incident wave, that is, $P_{\text{rad}} = A_m\mathcal{P}_{\text{inc}}$. It follows that:

$$A = eA_m \tag{15.3.2}$$

The effective area depends on the *direction* of arrival (θ,ϕ) of the incident wave. For all antennas, it can be shown that the effective area $A(\theta,\phi)$ is related to the power gain $G(\theta,\phi)$ and the wavelength $\lambda = c/f$ as follows:

$$\boxed{G(\theta,\phi) = \frac{4\pi A(\theta,\phi)}{\lambda^2}} \tag{15.3.3}$$

Similarly, because $G(\theta,\phi) = eD(\theta,\phi)$, the maximum effective aperture will be related to the directive gain by:

$$D(\theta,\phi) = \frac{4\pi A_m(\theta,\phi)}{\lambda^2} \tag{15.3.4}$$

In practice, the quoted effective area A of an antenna is the value corresponding to the direction of maximal gain $G_{\max}$. We write in this case:

$$\boxed{G_{\max} = \frac{4\pi A}{\lambda^2}} \tag{15.3.5}$$

Similarly, we have for the directivity $D_{\max} = 4\pi A_m/\lambda^2$. Because $D_{\max}$ is related to the beam solid angle by $D_{\max} = 4\pi/\Delta\Omega$, it follows that

$$D_{\max} = \frac{4\pi}{\Delta\Omega} = \frac{4\pi A_m}{\lambda^2} \quad\Rightarrow\quad A_m\Delta\Omega = \lambda^2 \tag{15.3.6}$$

Writing $\lambda = c/f$, we may express Eq. (15.3.5) in terms of frequency:

$$G_{\max} = \frac{4\pi f^2 A}{c^2} \tag{15.3.7}$$

The effective area is not equal to the physical area of an antenna. For example, linear antennas do not even have any characteristic physical area. For dish or horn antennas, on the other hand, the effective area is typically a fraction of the physical area (about 55–65 percent for dishes and 60–80 percent for horns.) For example, if the dish has a diameter of d meters, then we have:

$$\boxed{A = e_a\frac{1}{4}\pi d^2} \qquad \text{(effective area of dish antenna)} \tag{15.3.8}$$

where e_a is the *aperture efficiency* factor, typically $e_a = 0.55$–0.65. Combining Eqs. (15.3.5) and (15.3.8), we obtain:

$$\boxed{G_{\max} = e_a\left(\frac{\pi d}{\lambda}\right)^2} \tag{15.3.9}$$

Antennas fall into two classes: *fixed-area antennas*, such as dish antennas, for which A is independent of frequency, and *fixed-gain antennas*, such as linear antennas, for which G is independent of frequency. For fixed-area antennas, the gain *increases* quadratically with f. For fixed-gain antennas, A *decreases* quadratically with f.

Example 15.3.1: Linear antennas are fixed-gain antennas. For example, we will see in Sec. 16.1 that the gains of a (lossless) Hertzian dipole, a halfwave dipole, and a monopole antenna are the constants:

$$G_{\text{hertz}} = 1.5,\quad G_{\text{dipole}} = 1.64,\quad G_{\text{monopole}} = 3.28$$

Eq. (15.3.5) gives the effective areas $A = G\lambda^2/4\pi$:

$$A_{\text{hertz}} = 0.1194\lambda^2,\quad A_{\text{dipole}} = 0.1305\lambda^2,\quad A_{\text{monopole}} = 0.2610\lambda^2$$

In all cases the effective area is proportional to λ^2 and decreases with f^2. In the case of the commonly used monopole antenna, the effective area is approximately equal to a rectangle of sides λ and $\lambda/4$, the latter being the physical length of the monopole. □

Example 15.3.2: Determine the gain in dB of a dish antenna of diameter of 0.5 m operating at a satellite downlink frequency of 4 GHz and having 60% aperture efficiency. Repeat if the downlink frequency is 11 GHz. Repeat if the diameter is doubled to 1 m.

Solution: The effective area and gain of a dish antenna with diameter d is:

$$A = e_a\frac{1}{4}\pi d^2 \quad\Rightarrow\quad G = \frac{4\pi A}{\lambda^2} = e_a\left(\frac{\pi d}{\lambda}\right)^2 = e_a\left(\frac{\pi f d}{c}\right)^2$$

The calculated gains G in absolute and dB units are in the four cases:

	$d = 0.5$ m	$d = 1$ m
$f = 4$ GHz	263 = 24 dB	1052 = 30 dB
$f = 11$ GHz	1990 = 33 dB	7960 = 39 dB

Doubling the diameter (or the frequency) increases the gain by 6 dB, or a factor of 4. Conversely, if a dish antenna is to have a desired gain G (for example, to achieve a desired beamwidth), the above equation can be solved for the required diameter d in terms of G and f. □

The beamwidth of a dish antenna can be estimated by combining the approximate expression (15.2.20) with (15.3.5) and (15.3.8). Assuming a lossless antenna with diameter d and 100% aperture efficiency, and taking Eq. (15.2.20) literally, we have:

$$G_{\max} = \frac{4\pi A}{\lambda^2} = \left(\frac{\pi d}{\lambda}\right)^2 = D_{\max} = \frac{16}{\Delta\theta_B^2}$$

Solving for $\Delta\theta_B$, we obtain the expression in radians and in degrees:

$$\Delta\theta_B = \frac{4}{\pi}\,\frac{\lambda}{d} = 1.27\,\frac{\lambda}{d}\,, \qquad \Delta\theta_B = 73^{\circ}\,\frac{\lambda}{d} \tag{15.3.10}$$

Thus, the beamwidth depends inversely on the antenna diameter. In practice, quick estimates of the 3-dB beamwidth in degrees are obtained by replacing Eq. (15.3.10) by the formula [1174]:

$$\boxed{\Delta\theta_B = 1.22\,\frac{\lambda}{d} = 70^{\circ}\,\frac{\lambda}{d}} \qquad \text{(3-dB beamwidth of dish antenna)} \tag{15.3.11}$$

The constant 70° represents only a rough approximation (other choices are in the range 65–75°.) Solving for the ratio $d/\lambda = 1.22/\Delta\theta_B$ (here, $\Delta\theta_B$ is in radians), we may express the maximal gain inversely with $\Delta\theta_B^2$ as follows:

$$G_{\max} = e_a\left(\frac{\pi d}{\lambda}\right)^2 = \frac{e_a \pi^2 (1.22)^2}{\Delta\theta_B^2}$$

For a typical aperture efficiency of 60%, this expression can be written in the following approximate form, with $\Delta\theta_B$ given in degrees:

$$\boxed{G_{\max} = \frac{30\,000}{\Delta\theta_B^2}} \tag{15.3.12}$$

Equations (15.3.11) and (15.3.12) must be viewed as approximate design guidelines, or rules of thumb [1174], for the beamwidth and gain of a dish antenna.

Example 15.3.3: For the 0.5-m antenna of the previous example, estimate its beamwidth for the two downlink frequencies of 4 GHz and 11 GHz.

The operating wavelengths are in the two cases: $\lambda = 7.5$ cm and $\lambda = 2.73$ cm. Using Eq. (15.3.11), we find $\Delta\theta_B = 10.5^{\circ}$ and $\Delta\theta_B = 3.8^{\circ}$. □

Example 15.3.4: A geostationary satellite at height of 36,000 km is required to have earth coverage. Using the approximate design equations, determine the gain in dB and the diameter of the satellite antenna for a downlink frequency of 4 GHz. Repeat for 11 GHz.

Solution: This problem was considered in Example 15.2.3. The beamwidth angle for earth coverage was found to be $\Delta\theta_B = 17.36^{\circ}$. From Eq. (15.3.11), we find:

$$d = \lambda\,\frac{70^{\circ}}{\Delta\theta_B} = 7.5\,\frac{70^{\circ}}{17.36^{\circ}} = 30 \text{ cm}$$

From Eq. (15.3.12), we find:

$$G = \frac{30\,000}{\Delta\theta_B^2} = \frac{30\,000}{17.36^2} = 100 = 20 \text{ dB}$$

For 11 GHz, we find $d = 11$ cm, and G remains the same. □

In Eqs. (15.2.20) and (15.3.12), we implicitly assumed that the radiation pattern was independent of the azimuthal angle ϕ. When the pattern is not azimuthally symmetric, we may define two orthogonal polar directions parametrized, say, by angles θ_1 and θ_2, as shown in Fig. 15.3.2.

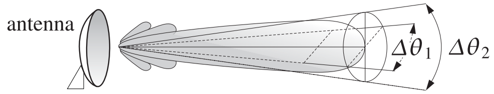


Fig. 15.3.2 Half-power beamwidths in two principal polar directions.

In this case $d\Omega = d\theta_1\,d\theta_2$ and we may approximate the beam solid angle by the product of the corresponding 3-dB beamwidths in these two directions, $\Delta\Omega = \Delta\theta_1\,\Delta\theta_2$. Then, the directivity takes the form (with the angles in radians and in degrees):

$$D_{\max} = \frac{4\pi}{\Delta\Omega} = \frac{4\pi}{\Delta\theta_1\,\Delta\theta_2} = \frac{41\,253}{\Delta\theta_1^{\circ}\,\Delta\theta_2^{\circ}} \tag{15.3.13}$$

Equations (15.3.12) and (15.3.13) are examples of a more general expression that relates directivity or gain to the 3-dB beamwidths for aperture antennas [1052,1064]:

$$\boxed{G_{\max} = \frac{p}{\Delta\theta_1\,\Delta\theta_2}} \tag{15.3.14}$$

where p is a *gain-beamwidth constant* whose value depends on the particular aperture antenna. We will see several examples of this relationship in Chapters 17 and 18. Practical values of p fall in the range 25 000–35 000 (with the beamwidth angles in degrees.)

15.4 Antenna Equivalent Circuits

To a generator feeding a transmitting antenna as in Fig. 15.2.1, the antenna appears as a load. Similarly, a receiver connected to a receiving antenna's output terminals will appear to the antenna as a load impedance. Such simple equivalent circuit representations of transmitting and receiving antennas are shown in Fig. 15.4.1, where in both cases V is the equivalent open-circuit Thévenin voltage.

In the transmitting antenna case, the antenna is represented by a load impedance Z_A, which in general will have both a resistive and a reactive part, $Z_A = R_A + jX_A$. The reactive part represents energy stored in the fields near the antenna, whereas the resistive part represents the power losses which arise because (a) power is radiated away from the antenna and (b) power is lost into heat in the antenna circuits and in the medium surrounding the antenna.

The generator has its own internal impedance $Z_G = R_G + jX_G$. The current at the antenna input terminals will be $I_{\text{in}} = V/(Z_G + Z_A)$, which allows us to determine (a) the total power P_{tot} produced by the generator, (b) the power P_T delivered to the antenna terminals, and (c) the power P_G lost in the generator's internal resistance R_G. These are:

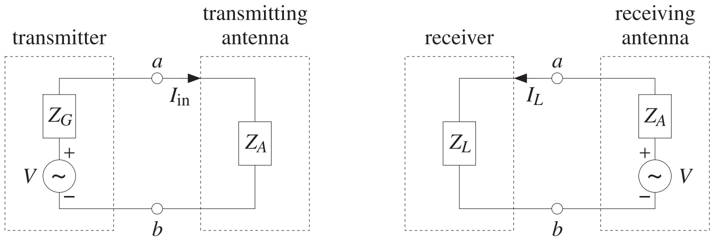


Fig. 15.4.1 Circuit equivalents of transmitting and receiving antennas.

$$
\begin{gathered}
P_{\text{tot}} = \frac{1}{2}\text{Re}(VI_{\text{in}}^*) = \frac{1}{2}\frac{|V|^2(R_G+R_A)}{|Z_G+Z_A|^2} \\
P_T = \frac{1}{2}|I_{\text{in}}|^2R_A = \frac{1}{2}\frac{|V|^2R_A}{|Z_G+Z_A|^2}\,, \quad P_G = \frac{1}{2}|I_{\text{in}}|^2R_G = \frac{1}{2}\frac{|V|^2R_G}{|Z_G+Z_A|^2}
\end{gathered}
\tag{15.4.1}
$$

It is evident that $P_{\text{tot}} = P_T + P_G$. A portion of the power P_T delivered to the antenna is radiated away, say an amount P_{rad}, and the rest is dissipated as ohmic losses, say P_{ohm}. Thus, $P_T = P_{\text{rad}} + P_{\text{ohm}}$. These two parts can be represented conveniently by equivalent resistances by writing $R_A = R_{\text{rad}} + R_{\text{ohm}}$, where R_{rad} is referred to as the *radiation resistance.* Thus, we have,

$$
P_T = \frac{1}{2}|I_{\text{in}}|^2R_A = \frac{1}{2}|I_{\text{in}}|^2R_{\text{rad}} + \frac{1}{2}|I_{\text{in}}|^2R_{\text{ohm}} = P_{\text{rad}} + P_{\text{ohm}}
\tag{15.4.2}
$$

The efficiency factor of Eq. (15.2.7) is evidently:

$$
e = \frac{P_{\text{rad}}}{P_T} = \frac{R_{\text{rad}}}{R_A} = \frac{R_{\text{rad}}}{R_{\text{rad}}+R_{\text{ohm}}}
$$

To maximize the amount of power P_T delivered to the antenna (and thus minimize the power lost in the generator's internal resistance), the load impedance must satisfy the usual *conjugate matching* condition:

$$
Z_A = Z_G^* \quad \Leftrightarrow \quad R_A = R_G\,,\quad X_A = -X_G
$$

In this case, $|Z_G + Z_A|^2 = (R_G+R_A)^2 + (X_G+X_A)^2 = 4R_G^2$, and it follows that the maximum power transferred to the load will be one-half the total—the other half being lost in R_G, that is,

$$
\boxed{P_{T,\max} = \frac{1}{2}P_{\text{tot}} = \frac{|V|^2}{8R_G}}
\tag{15.4.3}
$$

In the notation of Chap. 13, this is the *available* power from the generator. If the generator and antenna are mismatched, we have:

$$
P_T = \frac{|V|^2}{8R_G}\frac{4R_AR_G}{|Z_A+Z_G|^2} = P_{T,\max}\big(1-|\Gamma_{\text{gen}}|^2\big)\,, \qquad \Gamma_{\text{gen}} = \frac{Z_A-Z_G^*}{Z_A+Z_G}
\tag{15.4.4}
$$

Eq. (15.4.3) is often written in terms of the rms value of the source, that is, $V_{\text{rms}} = |V|/\sqrt{2}$, which leads to $P_{T,\max} = V_{\text{rms}}^2/4R_G$.

The case of a receiving antenna is similar. The induced currents on the antenna can be represented by a Thévenin-equivalent generator (the open-circuit voltage at the antenna output terminals) and an internal impedance Z_A. A consequence of the reciprocity principle is that Z_A is the same whether the antenna is transmitting or receiving.

The current into the load is $I_L = V/(Z_A + Z_L)$, where the load impedance is $Z_L = R_L + jX_L$. As before, we can determine the total power P_{tot} produced by the generator (i.e., intercepted by the antenna) and the power P_R delivered to the receiving load:

$$
P_{\text{tot}} = \frac{1}{2}\text{Re}(VI_L^*) = \frac{1}{2}\frac{|V|^2(R_L+R_A)}{|Z_L+Z_A|^2}\,, \quad P_R = \frac{1}{2}|I_L|^2R_L = \frac{1}{2}\frac{|V|^2R_L}{|Z_L+Z_A|^2}
\tag{15.4.5}
$$

Under conjugate matching, $Z_L = Z_A^*$, we find the maximum power delivered to the load:

$$
\boxed{P_{R,\max} = \frac{|V|^2}{8R_A}}
\tag{15.4.6}
$$

If the load and antenna are mismatched, we have:

$$
P_R = \frac{|V|^2}{8R_A}\frac{4R_AR_L}{|Z_L+Z_A|^2} = P_{R,\max}\big(1-|\Gamma_{\text{load}}|^2\big)\,, \qquad \Gamma_{\text{load}} = \frac{Z_L-Z_A^*}{Z_L+Z_A}
\tag{15.4.7}
$$

It is tempting to interpret the power dissipated in the internal impedance of the Thévenin circuit of the receiving antenna (that is, in Z_A) as representing the amount of power re-radiated or scattered by the antenna. However, with the exception of the so-called *minimum-scattering antennas*, such interpretation is not correct.

The issue has been discussed by Silver [21] and more recently in Refs. [1028–1031]. See also Refs. [1004–1027] for further discussion of the transmitting, receiving, and scattering properties of antennas.

15.5 Effective Length

The polarization properties of the electric field radiated by an antenna depend on the transverse component of the radiation vector $\boldsymbol{F}_\perp$ according to Eq. (14.10.5):

$$
\boldsymbol{E} = -jk\eta\,\frac{e^{-jkr}}{4\pi r}\,\boldsymbol{F}_\perp = -jk\eta\,\frac{e^{-jkr}}{4\pi r}\,(F_\theta\,\hat{\boldsymbol{\theta}} + F_\phi\,\hat{\boldsymbol{\phi}})
$$

The vector *effective length*, or *effective height* of a transmitting antenna is defined in terms of $\boldsymbol{F}_\perp$ and the input current to the antenna terminals I_{in} as follows [996]:[†]

$$
\boldsymbol{h} = -\frac{\boldsymbol{F}_\perp}{I_{\text{in}}} \qquad \text{(effective length)}
\tag{15.5.1}
$$

In general, $\boldsymbol{h}$ is a function of θ, ϕ. The electric field is, then, written as:

$$
\boldsymbol{E} = jk\eta\,\frac{e^{-jkr}}{4\pi r}\,I_{\text{in}}\,\boldsymbol{h}
\tag{15.5.2}
$$

[†] Often, it is defined with a positive sign $\boldsymbol{h} = \boldsymbol{F}_\perp/I_{\text{in}}$.

The definition of $\boldsymbol{h}$ is motivated by the case of a z-directed Hertzian dipole antenna, which can be shown to have $\boldsymbol{h} = l\sin\theta\,\hat{\boldsymbol{\theta}}$. More generally, for a z-directed linear antenna with current $I(z)$, it follows from Eq. (16.1.5) that:

$$\boldsymbol{h}(\theta)= h(\theta)\,\hat{\boldsymbol{\theta}}\,, \qquad h(\theta)= \sin\theta\,\frac{1}{I_{\text{in}}}\int_{-l/2}^{l/2} I(z')\,e^{jkz'\cos\theta}dz' \tag{15.5.3}$$

As a consequence of the reciprocity principle, it can be shown [996] that the *open-circuit voltage* V at the terminals of a receiving antenna is given in terms of the effective length and the incident field $\boldsymbol{E}_i$ by:

$$V = \boldsymbol{E}_i\cdot\boldsymbol{h} \tag{15.5.4}$$

The normal definition of the effective area of an antenna and the result $G = 4\pi A/\lambda^2$ depend on the assumptions that the receiving antenna is conjugate-matched to its load and that the polarization of the incident wave matches that of the antenna.

The effective length helps to characterize the degree of polarization mismatch that may exist between the incident field and the antenna. To see how the gain-area relationship must be modified, we start with the definition (15.3.1) and use (15.4.5):

$$A(\theta,\phi)= \frac{P_R}{\mathcal{P}_{\text{inc}}} = \frac{\dfrac{1}{2}R_L|I_L|^2}{\dfrac{1}{2\eta}|\boldsymbol{E}_i|^2} = \frac{\eta R_L|V|^2}{|Z_L+Z_A|^2|\boldsymbol{E}_i|^2} = \frac{\eta R_L|\boldsymbol{E}_i\cdot\boldsymbol{h}|^2}{|Z_L+Z_A|^2|\boldsymbol{E}_i|^2}$$

Next, we define the polarization and load *mismatch factors* by:

$$\begin{aligned} e_{\text{pol}} &= \frac{|\boldsymbol{E}_i\cdot\boldsymbol{h}|^2}{|\boldsymbol{E}_i|^2\,|\boldsymbol{h}|^2} \\ e_{\text{load}} &= \frac{4R_LR_A}{|Z_L+Z_A|^2} = 1-|\Gamma_{\text{load}}|^2\,, \quad \text{where} \quad \Gamma_{\text{load}} = \frac{Z_L-Z_A^*}{Z_L+Z_A} \end{aligned} \tag{15.5.5}$$

The effective area can be written then in the form:

$$A(\theta,\phi)= \frac{\eta|\boldsymbol{h}|^2}{4R_A}\,e_{\text{load}}\,e_{\text{pol}} \tag{15.5.6}$$

On the other hand, using (15.1.4) and (15.4.1), the power gain may be written as:

$$G(\theta,\phi)= \frac{4\pi U(\theta,\phi)}{P_T} = \frac{4\pi\dfrac{\eta k^2|\boldsymbol{F}_\perp|^2}{32\pi^2}}{\dfrac{1}{2}R_A|I_{\text{in}}|^2} = \frac{\pi\eta|\boldsymbol{h}|^2}{\lambda^2R_A} \quad\Rightarrow\quad \frac{\eta|\boldsymbol{h}|^2}{4R_A} = \frac{\lambda^2}{4\pi}G(\theta,\phi)$$

Inserting this in Eq. (15.5.6), we obtain the modified area-gain relationship [997]:

$$\boxed{A(\theta,\phi)= e_{\text{load}}\,e_{\text{pol}}\,\frac{\lambda^2}{4\pi}G(\theta,\phi)} \tag{15.5.7}$$

Assuming that the incident field originates at some antenna with its own effective length $\boldsymbol{h}_i$, then $\boldsymbol{E}_i$ will be proportional to $\boldsymbol{h}_i$ and we may write the polarization mismatch factor in the following form:

$$e_{\text{pol}} = \frac{|\boldsymbol{h}_i\cdot\boldsymbol{h}|^2}{|\boldsymbol{h}_i|^2\,|\boldsymbol{h}|^2} = |\hat{\boldsymbol{h}}_i\cdot\hat{\boldsymbol{h}}|^2\,, \qquad \text{where} \quad \hat{\boldsymbol{h}}_i = \frac{\boldsymbol{h}_i}{|\boldsymbol{h}_i|}\,, \quad \hat{\boldsymbol{h}} = \frac{\boldsymbol{h}}{|\boldsymbol{h}|}$$

When the load is conjugate-matched, we have $e_{\text{load}} = 1$, and when the incident field has matching polarization with the antenna, that is, $\hat{\boldsymbol{h}}_i = \hat{\boldsymbol{h}}^*$, then, $e_{\text{pol}} = 1$.

15.6 Communicating Antennas

The communication between a transmitting and a receiving antenna can be analyzed with the help of the concept of gain and effective area. Consider two antennas oriented towards the maximal gain of each other and separated by a distance r, as shown in Fig. 15.6.1.

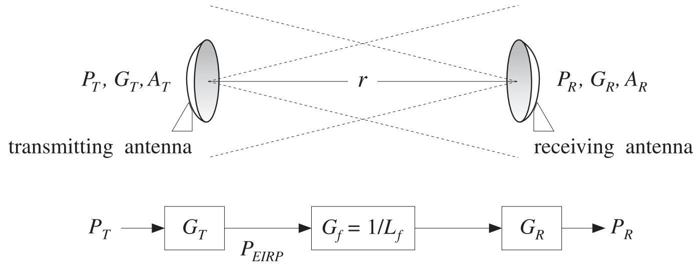


Fig. 15.6.1 Transmitting and receiving antennas.

Let $\{P_T, G_T, A_T\}$ be the power, gain, and effective area of the transmitting antenna, and $\{P_R, G_R, A_R\}$ be the same quantities for the receiving antenna. In the direction of the receiving antenna, the transmitting antenna has $P_{\text{EIRP}} = P_TG_T$ and establishes a power density at distance r:

$$\mathcal{P}_T = \frac{dP_T}{dS} = \frac{P_{\text{EIRP}}}{4\pi r^2} = \frac{P_TG_T}{4\pi r^2} \tag{15.6.1}$$

From the incident power density $\mathcal{P}_T$, the receiving antenna extracts power P_R given in terms of the effective area A_R as follows:

$$\boxed{P_R = A_R\mathcal{P}_T = \frac{P_TG_TA_R}{4\pi r^2}} \qquad \text{(Friis formula)} \tag{15.6.2}$$

This is known as the *Friis formula* for communicating antennas and can be written in several different equivalent forms. Replacing G_T in terms of the transmitting antenna's effective area A_T, that is, $G_T = 4\pi A_T/\lambda^2$, Eq. (15.6.2) becomes:

$$P_R = \frac{P_TA_TA_R}{\lambda^2r^2} \tag{15.6.3}$$

A better way of rewriting Eq. (15.6.2) is as a product of gain factors. Replacing $A_R = \lambda^2G_R/4\pi$, we obtain:

$$P_R = \frac{P_TG_TG_R\lambda^2}{(4\pi r)^2} \tag{15.6.4}$$

The effect of the propagation path, which causes P_R to attenuate with the square of the distance r, can be quantified by defining the *free-space loss* and gain by

$$\boxed{L_f = \left(\frac{4\pi r}{\lambda}\right)^2, \quad G_f = \frac{1}{L_f} = \left(\frac{\lambda}{4\pi r}\right)^2} \qquad \text{(free-space loss and gain)} \tag{15.6.5}$$

Then, Eq. (15.6.4) can be written as the product of the transmit and receive gains and the propagation loss factor:

$$\boxed{P_R = P_T G_T\left(\frac{\lambda}{4\pi r}\right)^2 G_R = P_T G_T\,\frac{1}{L_f}\,G_R = P_T G_T G_f G_R} \tag{15.6.6}$$

Such a gain model for communicating antennas is illustrated in Fig. 15.6.1. An additional loss factor, $G_{\text{other}} = 1/L_{\text{other}}$, may be introduced, if necessary, representing other losses, such as atmospheric absorption and scattering. It is customary to express Eq. (15.6.6) additively in dB, where $(P_R)_{\text{dB}} = 10\log_{10}P_R$, $(G_T)_{\text{dB}} = 10\log_{10}G_T$, etc.:

$$(P_R)_{\text{dB}} = (P_T)_{\text{dB}} + (G_T)_{\text{dB}} - (L_f)_{\text{dB}} + (G_R)_{\text{dB}} \tag{15.6.7}$$

Example 15.6.1: A geosynchronous satellite is transmitting a TV signal to an earth-based station at a distance of 40,000 km. Assume that the dish antennas of the satellite and the earth station have diameters of 0.5 m and 5 m, and aperture efficiencies of 60%. If the satellite's transmitter power is 6 W and the downlink frequency 4 GHz, calculate the antenna gains in dB and the amount of received power.

Solution: The wavelength at 4 GHz is $\lambda = 7.5$ cm. The antenna gains are calculated by:

$$G = e_a\left(\frac{\pi d}{\lambda}\right)^2 \quad\Rightarrow\quad G_{\text{sat}} = 263.2 = 24\text{ dB}, \quad G_{\text{earth}} = 26320 = 44\text{ dB}$$

Because the ratio of the earth and satellite antenna diameters is 10, the corresponding gains will differ by a ratio of 100, or 20 dB. The satellite's transmitter power is in dB, $P_T = 10\log_{10}(6) = 8$ dBW, and the free-space loss and gain:

$$L_f = \left(\frac{4\pi r}{\lambda}\right)^2 = 4\times 10^{19} \quad\Rightarrow\quad L_f = 196\text{ dB}, \quad G_f = -196\text{ dB}$$

It follows that the received power will be in dB:

$$P_R = P_T + G_T - L_f + G_R = 8 + 24 - 196 + 44 = -120\text{ dBW} \quad\Rightarrow\quad P_R = 10^{-12}\text{ W}$$

or, $P_R = 1$ pW (pico-watt). Thus, the received power is extremely small. □

When the two antennas are mismatched in their polarization with a mismatch factor $e_{\text{pol}} = |\hat{\boldsymbol{h}}_R\cdot\hat{\boldsymbol{h}}_T|^2$, and the receiving antenna is mismatched to its load with $e_{\text{load}} = 1-|\Gamma_{\text{load}}|^2$, then the Friis formula (15.6.2) is still valid, but replacing A_R using Eq. (15.5.7), leads to a modified form of Eq. (15.6.4):

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi r)^2}\,|\hat{\boldsymbol{h}}_R\cdot\hat{\boldsymbol{h}}_T|^2\,(1-|\Gamma_{\text{load}}|^2) \tag{15.6.8}$$

15.7 Antenna Noise Temperature

We saw in the above example that the received signal from a geosynchronous satellite is extremely weak, of the order of picowatts, because of the large free-space loss which is typically of the order of 200 dB.

To be able to detect such a weak signal, the receiving system must maintain a noise level that is lower than the received signal. Noise is introduced into the receiving system by several sources.

In addition to the desired signal, the receiving antenna picks up noisy signals from the sky, the ground, the weather, and other natural or man-made noise sources. These noise signals, coming from different directions, are weighted according to the antenna gain and result into a weighted average noise power at the output terminals of the antenna. For example, if the antenna is pointing straight up into the sky, it will still pick up through its sidelobes some reflected signals as well as thermal noise from the ground. Ohmic losses in the antenna itself will be another source of noise.

The antenna output is sent over a feed line (such as a waveguide or transmission line) to the receiver circuits. The lossy feed line will attenuate the signal further and also introduce its own thermal noise.

The output of the feed line is then sent into a low-noise-amplifier (LNA), which preamplifies the signal and introduces only a small amount of thermal noise. The low-noise nature of the LNA is a critical property of the receiving system.

The output of the LNA is then passed on to the rest of the receiving system, consisting of downconverters, IF amplifiers, and so on. These subsystems will also introduce their own gain factors and thermal noise.

Such a cascade of receiver components is depicted in Fig. 15.7.1. The sum total of all the noises introduced by these components must be maintained at acceptably low levels (relative to the amplified desired signal.)

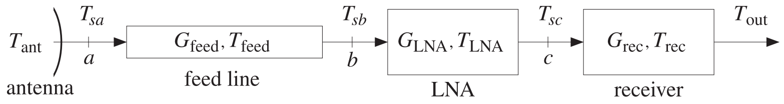

T_{ant}
antenna
T_{sa}
a
$G_{\text{feed}}, T_{\text{feed}}$
feed line
T_{sb}
b
$G_{\text{LNA}}, T_{\text{LNA}}$
LNA
T_{sc}
c
$G_{\text{rec}}, T_{\text{rec}}$
receiver
T_{out}

Fig. 15.7.1 Typical receiving antenna system.

The *average power* N (in Watts) of a noise source within a certain bandwidth of B Hz can be quantified by means of an *equivalent temperature* T defined through:

$$\boxed{N = kTB} \qquad \text{(noise power within bandwidth } B\text{)} \tag{15.7.1}$$

where k is Boltzmann's constant $k = 1.3803\times 10^{-23}$ W/Hz K and T is in degrees Kelvin. The temperature T is not necessarily the physical temperature of the source, it only provides a convenient way to express the noise power. (For a thermal source, T is indeed the physical temperature.) Eq. (15.7.1) is commonly expressed in dB as:

$$N_{\text{dB}} = T_{\text{dB}} + B_{\text{dB}} + k_{\text{dB}} \tag{15.7.2}$$

where $T_{\mathrm{dB}} = 10\log_{10} T$, $B_{\mathrm{dB}} = 10\log_{10} B$, and $k_{\mathrm{dB}} = 10\log_{10} k$ is Boltzmann's constant in dB: $k_{\mathrm{dB}} = -228.6$ dB. Somewhat incorrectly, but very suggestively, the following units are used in practice for the various terms in Eq. (15.7.2):

$$\mathrm{dB\,W} = \mathrm{dB\,K} + \mathrm{dB\,Hz} + \mathrm{dB\,W/Hz\,K}$$

The bandwidth B depends on the application. For example, satellite transmissions of TV signals require a bandwidth of about 30 MHz. Terrestrial microwave links may have B of 60 MHz. Cellular systems may have B of the order of 30 kHz for AMPS or 200 kHz for GSM.

Example 15.7.1: Assuming a 30-MHz bandwidth, we give below some examples of noise powers and temperatures and compute the corresponding signal-to-noise ratio S/N, relative to a 1 pW reference signal ($S = 1$ pW).

T	T_{dB}	$N = kTB$	N_{dB}	S/N
50 K	17.0 dBK	0.0207 pW	−136.8 dBW	16.8 dB
100 K	20.0 dBK	0.0414 pW	−133.8 dBW	13.8 dB
200 K	23.0 dBK	0.0828 pW	−130.8 dBW	10.8 dB
290 K	24.6 dBK	0.1201 pW	−129.2 dBW	9.2 dB
500 K	27.0 dBK	0.2070 pW	−126.8 dBW	6.8 dB
1000 K	30.0 dBK	0.4141 pW	−123.8 dBW	3.8 dB
2400 K	33.8 dBK	1.0000 pW	−120.0 dBW	0.0 dB

The last example shows that 2400 K corresponds to 1 pW noise. □

The average noise power N_{ant} at the antenna terminals is characterized by an equivalent *antenna noise temperature* T_{ant}, such that $N_{\mathrm{ant}} = kT_{\mathrm{ant}}B$.

The temperature T_{ant} represents the weighted contributions of all the radiating noise sources picked up by the antenna through its mainlobe and sidelobes. The value of T_{ant} depends primarily on the orientation and elevation angle of the antenna, and what the antenna is looking at.

Example 15.7.2: An earth antenna looking at the sky "sees" a noise temperature T_{ant} of the order of 30–60 K. Of that, about 10 K arises from the mainlobe and sidelobes pointing towards the sky and 20–40 K from sidelobes pointing backwards towards the earth [5,1044–1048]. In rainy weather, T_{ant} might increase by 60 K or more.

The sky noise temperature depends on the elevation angle of the antenna. For example, at an elevations of 5°, 10°, and 30°, the sky temperature is about 20 K, 10 K, and 4 K at 4 GHz, and 25 K, 12 K, and 5 K at 6 GHz [1044]. □

Example 15.7.3: The noise temperature of the earth viewed from space, such as from a satellite, is about 254 K. This is obtained by equating the sun's energy that is absorbed by the earth to the thermal radiation from the earth [1044]. □

Example 15.7.4: For a base station cellular antenna looking horizontally, atmospheric noise temperature contributes about 70–100 K at the cellular frequency of 1 GHz, and man-made noise contributes another 10–120 K depending on the area (rural or urban). The total value of T_{ant} for cellular systems is in the range of 100–200 K [1049,1050]. □

In general, a noise source in some direction (θ,ϕ) will be characterized by an effective noise temperature $T(\theta,\phi)$, known as the *brightness temperature*, such that the radiated noise power in that direction will be $N(\theta,\phi) = kT(\theta,\phi)B$. The antenna temperature T_{ant} will be given by the average over all such sources weighted by the receiving gain of the antenna:

$$\boxed{T_{\mathrm{ant}} = \frac{1}{\Delta\Omega}\int_0^{\pi}\int_0^{2\pi} T(\theta,\phi)g(\theta,\phi)\,d\Omega} \tag{15.7.3}$$

where $\Delta\Omega$ is the beam solid angle of the antenna. It follows from Eq. (15.2.15) that $\Delta\Omega$ serves as a normalization factor for this average:

$$\Delta\Omega = \int_0^{\pi}\int_0^{2\pi} g(\theta,\phi)\,d\Omega \tag{15.7.4}$$

Eq. (15.7.3) can also be written in the following equivalent forms, in terms of the directive gain or the effective area of the antenna:

$$T_{\mathrm{ant}} = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} T(\theta,\phi)D(\theta,\phi)\,d\Omega = \frac{1}{\lambda^2}\int_0^{\pi}\int_0^{2\pi} T(\theta,\phi)A(\theta,\phi)\,d\Omega$$

As an example of Eq. (15.7.3), we consider the case of a *point* source, such as the sun, the moon, a planet, or a radio star. Then, Eq. (15.7.3) gives:

$$T_{\mathrm{ant}} = T_{\mathrm{point}}\frac{g_{\mathrm{point}}\Delta\Omega_{\mathrm{point}}}{\Delta\Omega}$$

where g_{point} and $\Delta\Omega_{\mathrm{point}}$ are the antenna gain in the direction of the source and the small solid angle subtended by the source. If the antenna's mainlobe is pointing towards that source then, $g_{\mathrm{point}} = 1$.

As another example, consider the case of a spatially *extended* noise source, such as the sky, which is assumed to have a constant temperature T_{ext} over its angular width. Then, Eq. (15.7.3) becomes:

$$T_{\mathrm{ant}} = T_{\mathrm{ext}}\frac{\Delta\Omega_{\mathrm{ext}}}{\Delta\Omega}\,, \qquad \text{where} \quad \Delta\Omega_{\mathrm{ext}} = \int_{\mathrm{ext}} g(\theta,\phi)\,d\Omega$$

The quantity $\Delta\Omega_{\mathrm{ext}}$ is the portion of the antenna's beam solid angle occupied by the extended source.

As a third example, consider the case of an antenna pointing towards the sky and picking up the atmospheric sky noise through its mainlobe and partly through its sidelobes, and also picking up noise from the ground through the rest of its sidelobes. Assuming the sky and ground noise temperatures are uniform over their spatial extents, Eq. (15.7.3) will give approximately:

$$T_{\mathrm{ant}} = T_{\mathrm{sky}}\frac{\Delta\Omega_{\mathrm{sky}}}{\Delta\Omega} + T_{\mathrm{ground}}\frac{\Delta\Omega_{\mathrm{ground}}}{\Delta\Omega}$$

where $\Delta\Omega_{\mathrm{sky}}$ and $\Delta\Omega_{\mathrm{ground}}$ are the portions of the beam solid angle occupied by the sky and ground:

$$\Delta\Omega_{\mathrm{sky}} = \int_{\mathrm{sky}} g(\theta,\phi)\,d\Omega\,, \qquad \Delta\Omega_{\mathrm{ground}} = \int_{\mathrm{ground}} g(\theta,\phi)\,d\Omega$$

Assuming that the sky and ground beam solid angles account for the total beam solid angle, we have

$$\Delta\Omega = \Delta\Omega_{\text{sky}} + \Delta\Omega_{\text{ground}}$$

The sky and ground *beam efficiency* ratios may be defined by:

$$e_{\text{sky}} = \frac{\Delta\Omega_{\text{sky}}}{\Delta\Omega}\,, \qquad e_{\text{ground}} = \frac{\Delta\Omega_{\text{ground}}}{\Delta\Omega}\,, \qquad e_{\text{sky}} + e_{\text{ground}} = 1$$

Then, the antenna noise temperature can be written in the form:

$$\boxed{T_{\text{ant}} = e_{\text{sky}}T_{\text{sky}} + e_{\text{ground}}T_{\text{ground}}} \tag{15.7.5}$$

Example 15.7.5: At 4 GHz and elevation angle of 30°, the sky noise temperature is about 4 K. Assuming a ground temperature of 290 K and that 90% of the beam solid angle of an earth-based antenna is pointing towards the sky and 10% towards the ground, we calculate the effective antenna temperature:

$$T_{\text{ant}} = e_{\text{sky}}T_{\text{sky}} + e_{\text{ground}}T_{\text{ground}} = 0.9\times 4 + 0.1\times 290 = 32.6\text{ K}$$

If the beam efficiency towards the sky changes to 85%, then $e_{\text{sky}} = 0.85$, $e_{\text{ground}} = 0.15$ and we find $T_{\text{ant}} = 46.9$ K. □

The mainlobe and sidelobe *beam efficiencies* of an antenna represent the proportions of the beam solid angle occupied by the mainlobe and sidelobe of the antenna. The corresponding beam solid angles are defined by:

$$\Delta\Omega = \int_{\text{tot}} g(\theta,\phi)\,d\Omega = \int_{\text{main}} g(\theta,\phi)\,d\Omega + \int_{\text{side}} g(\theta,\phi)\,d\Omega = \Delta\Omega_{\text{main}} + \Delta\Omega_{\text{side}}$$

Thus, the beam efficiencies will be:

$$e_{\text{main}} = \frac{\Delta\Omega_{\text{main}}}{\Delta\Omega}\,, \qquad e_{\text{side}} = \frac{\Delta\Omega_{\text{side}}}{\Delta\Omega}\,, \qquad e_{\text{main}} + e_{\text{side}} = 1$$

Assuming that the entire mainlobe and a fraction, say α, of the sidelobes point towards the sky, and therefore, a fraction $(1-\alpha)$ of the sidelobes will point towards the ground, we may express the sky and ground beam solid angles as follows:

$$\begin{aligned}\Delta\Omega_{\text{sky}} &= \Delta\Omega_{\text{main}} + \alpha\Delta\Omega_{\text{side}}\\ \Delta\Omega_{\text{ground}} &= (1-\alpha)\Delta\Omega_{\text{side}}\end{aligned}$$

or, in terms of the efficiency factors:

$$\begin{aligned}e_{\text{sky}} &= e_{\text{main}} + \alpha e_{\text{side}} = e_{\text{main}} + \alpha(1-e_{\text{main}})\\ e_{\text{ground}} &= (1-\alpha)e_{\text{side}} = (1-\alpha)(1-e_{\text{main}})\end{aligned}$$

Example 15.7.6: Assuming an 80% mainlobe beam efficiency and that half of the sidelobes point towards the sky and the other half towards the ground, we have $e_{\text{main}} = 0.8$ and $\alpha = 0.5$, which lead to the sky beam efficiency $e_{\text{sky}} = 0.9$. □

15.8 *System Noise Temperature*

In a receiving antenna system, the signal-to-noise ratio at the receiver must take into account not only the noise picked up by the antenna, and quantified by T_{ant}, but also all the internal noises introduced by the various components of the receiver.

Every device, passive or active, is a source of noise generated internally. Such noise may be modeled as an internal noise source acting at the input of the device, as shown in Fig. 15.8.1. (Alternatively, the noise source can be added at the output, but the input convention is more common.)

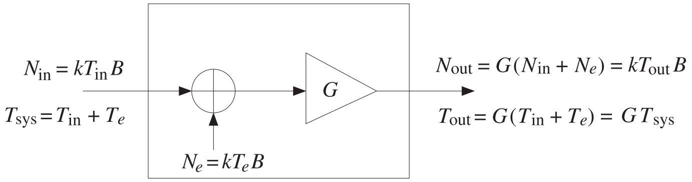

$N_{\text{in}} = kT_{\text{in}}B$
$T_{\text{sys}} = T_{\text{in}} + T_e$
$N_e = kT_eB$
G
$N_{\text{out}} = G(N_{\text{in}} + N_e) = kT_{\text{out}}B$
$T_{\text{out}} = G(T_{\text{in}} + T_e) = GT_{\text{sys}}$

Fig. 15.8.1 Noise model of a device.

The amount of added noise power is expressed in terms of the *effective noise temperature* T_e of the device:

$$\boxed{N_e = kT_eB} \qquad \text{(effective internal noise)} \tag{15.8.1}$$

The sum of N_e and the noise power of the input signal N_{in} will be the total noise power, or the *system noise power* at the input to the device. If the input noise is expressed in terms of its own noise temperature, $N_{\text{in}} = kT_{\text{in}}B$, we will have:

$$N_{\text{sys}} = N_{\text{in}} + N_e = k(T_{\text{in}} + T_e)B = kT_{\text{sys}}B \qquad \text{(total input noise)} \tag{15.8.2}$$

where we introduced the *system noise temperature*[†] at the device input:

$$\boxed{T_{\text{sys}} = T_{\text{in}} + T_e} \qquad \text{(system noise temperature)} \tag{15.8.3}$$

If the device has power gain G,[‡] then the noise power at the output of the device and its equivalent temperature, $N_{\text{out}} = kT_{\text{out}}B$, can be expressed as follows:

$$\begin{aligned}N_{\text{out}} &= G(N_{\text{in}} + N_e) = GN_{\text{sys}}\\ T_{\text{out}} &= G(T_{\text{in}} + T_e) = GT_{\text{sys}}\end{aligned} \tag{15.8.4}$$

One interpretation of the system noise power $N_{\text{sys}} = kT_{\text{sys}}B$ is that it represents the required input power to an equivalent *noiseless* system (with the same gain) that will produce the same output power as the actual noisy system.

If a desired signal with noise power S_{in} is also input to the device, then the signal power at the output will be $S_{\text{out}} = GS_{\text{in}}$. The *system signal-to-noise ratio* is defined to be the ratio of the input signal power to total system noise power:

[†] Also called the *operating noise temperature*.

[‡] More precisely, G is the *available* power gain of the device, in the notation of Sec. 13.6.

$$\boxed{SNR_{\text{sys}} = \frac{S_{\text{in}}}{N_{\text{sys}}} = \frac{S_{\text{in}}}{kT_{\text{sys}}B} = \frac{S_{\text{in}}}{k(T_{\text{in}} + T_e)B}} \qquad \text{(system SNR)} \tag{15.8.5}$$

The SNR is the same whether it is measured at the input or the output of the device; indeed, multiplying numerator and denominator by G and using (15.8.4), we have:

$$SNR_{\text{sys}} = \frac{S_{\text{in}}}{N_{\text{sys}}} = \frac{S_{\text{out}}}{N_{\text{out}}} \tag{15.8.6}$$

A related concept is that of the *noise figure* of the device, which also characterizes the internally generated noise. It is related to the effective noise temperature T_e by:

$$F = 1 + \frac{T_e}{T_0} \quad \Leftrightarrow \quad T_e = (F-1)T_0 \tag{15.8.7}$$

where T_0 is the standardized constant temperature $T_0 = 290\,\text{K}$.

The device of Fig. 15.8.1 can be passive or active. The case of a *passive attenuator*, such as a lossy transmission line or waveguide connecting the antenna to the receiver, deserves special treatment.

In this case, the gain G will be less than unity $G < 1$, representing a power loss. For a line of length l and attenuation constant α (nepers per meter), we will have $G = e^{-2\alpha l}$. The corresponding loss factor will be $L = G^{-1} = e^{2\alpha l}$. If $\alpha l \ll 1$, we can write approximately $G = 1 - 2\alpha l$ and $L = 1 + 2\alpha l$.

If the physical temperature of the line is T_{phys} then, from either the input or output end, the line will appear as a thermal noise source of power $kT_{\text{phys}}B$. Therefore, the condition $N_{\text{in}} = N_{\text{out}} = kT_{\text{phys}}B$ implies that $kT_{\text{phys}}B = Gk(T_{\text{phys}} + T_e)B$, which gives:

$$\boxed{T_e = \frac{1}{G}(1-G)T_{\text{phys}} = (L-1)T_{\text{phys}}} \qquad \text{(attenuator)} \tag{15.8.8}$$

If the physical temperature is $T_{\text{phys}} = T_0 = 290\,\text{K}$, then, by comparing to Eq. (15.8.7) it follows that the noise figure of the attenuator will be equal to its loss:

$$T_e = (L-1)T_0 = (F-1)T_0 \quad \Rightarrow \quad F = L = \frac{1}{G}$$

When the input to the attenuator is an external noise source of power $N_{\text{in}} = kT_{\text{in}}B$, the system noise temperature at the input and at the output of the attenuator will be:

$$\begin{aligned} T_{\text{sys}} &= T_{\text{in}} + T_e = T_{\text{in}} + (L-1)T_{\text{phys}} \\ T_{\text{out}} &= GT_{\text{sys}} = GT_{\text{in}} + (1-G)T_{\text{phys}} = \frac{1}{L}T_{\text{in}} + \left(1 - \frac{1}{L}\right)T_{\text{phys}} \end{aligned} \tag{15.8.9}$$

The last equation can be expressed in terms of the input and output powers $N_{\text{in}} = kT_{\text{in}}B$ and $N_{\text{out}} = kT_{\text{out}}B$:

$$N_{\text{out}} = \frac{1}{L}N_{\text{in}} + \left(1 - \frac{1}{L}\right)kT_{\text{phys}}B \tag{15.8.10}$$

Thus, the input power is attenuated as expected, but the attenuator also adds its own thermal noise power. More generally, if the input power arises from signal plus noise $P_{\text{in}} = S_{\text{in}} + N_{\text{in}}$, the power at the output will be $P_{\text{out}} = S_{\text{out}} + N_{\text{out}} = GS_{\text{in}} + N_{\text{out}}$:

$$P_{\text{out}} = \frac{1}{L}P_{\text{in}} + \left(1 - \frac{1}{L}\right)kT_{\text{phys}}B \tag{15.8.11}$$

When two or more devices are cascaded, each will contribute its own internal noise. Fig. 15.8.2 shows two such devices with available power gains G_1 and G_2 and effective noise temperatures T_1 and T_2. The cascade combination can be replaced by an *equivalent* device with gain G_1G_2 and effective noise temperature T_{12}.

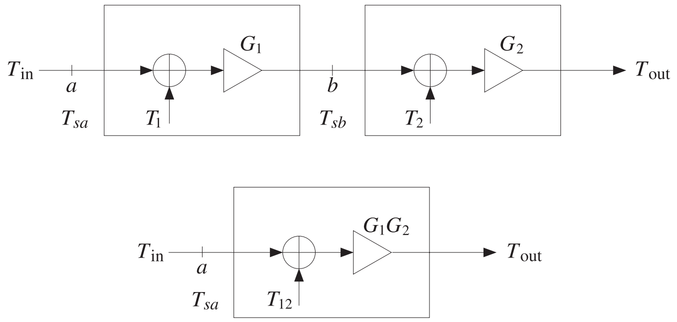


Fig. 15.8.2 Equivalent noise model of two cascaded devices.

The equivalent temperature T_{12} can be determined by superposition. The internal noise power added by the first device, $N_1 = kT_1B$, will go through the gains G_1 and G_2 and will contribute an amount $G_1G_2N_1$ to the output. The noise generated by the second device, $N_2 = kT_2B$, will contribute an amount G_2N_2. The sum of these two powers will be equivalent to the amount contributed to the output by the combined system, $G_1G_2N_{12} = G_1G_2kT_{12}B$. Thus,

$$G_1G_2kT_{12}B = G_1G_2kT_1B + G_2kT_2B \quad \Rightarrow \quad G_1G_2T_{12} = G_1G_2T_1 + G_2T_2$$

It follows that:

$$\boxed{T_{12} = T_1 + \frac{1}{G_1}T_2} \qquad \text{(equivalent noise temperature)} \tag{15.8.12}$$

If G_1 is a large gain, $G_1 \gg 1$, then the contribution of the second device is reduced drastically. On the other hand, if the first device is an attenuator, such as a transmission line, then the contribution of T_2 will be amplified because $G_1 < 1$.

According to Eqs. (15.8.3) and (15.8.4), the system noise temperatures at the overall input, at the output of G_1, and at the overall output will be:

$$\begin{aligned} T_{\text{sys}} &= T_{sa} = T_{\text{in}} + T_{12} = T_{\text{in}} + T_1 + \frac{1}{G_1}T_2 \\ T_{sb} &= G_1T_{sa} = G_1(T_{\text{in}} + T_1) + T_2 \\ T_{\text{out}} &= G_2T_{sb} = G_1G_2T_{\text{sys}} = G_1G_2(T_{\text{in}} + T_1) + G_2T_2 \end{aligned} \tag{15.8.13}$$

The system SNR will be:

$$SNR_{\text{sys}} = \frac{S_{\text{in}}}{kT_{\text{sys}}B} = \frac{S_{\text{in}}}{k(T_{\text{in}} + T_{12})B}$$

The signal powers at points *a*, *b*, and at the output will be $S_a = S_{\text{in}}$, $S_b = G_1 S_a$, and $S_{\text{out}} = G_2 S_b = G_1 G_2 S_a$. It follows from Eq. (15.8.13) that the system SNR will be the *same*, regardless of whether it is referred to the point *a*, the point *b*, or the overall output:

$$SNR_{\text{sys}} = SNR_a = SNR_b = SNR_{\text{out}}$$

For three cascaded devices, shown in Fig. 15.8.3, any pair of two consecutive ones can be replaced by its equivalent, according to Eq. (15.8.12). For example, the first two can be replaced by T_{12} and then combined with T_3 to give the overall equivalent temperature:

$$T_{12} = T_1 + \frac{1}{G_1}T_2\,, \qquad T_{123} = T_{12} + \frac{1}{G_1G_2}T_3$$

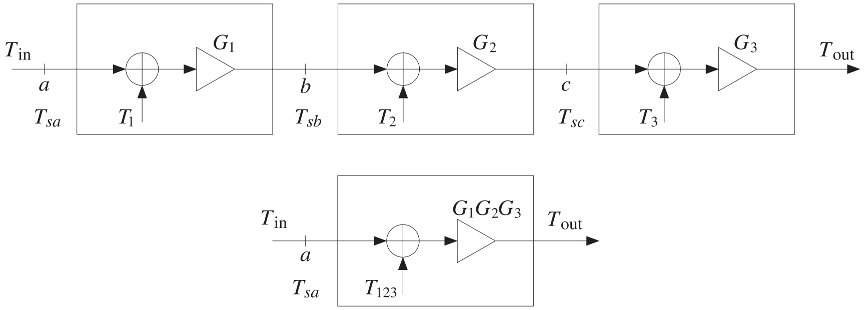


Fig. 15.8.3 Equivalent noise temperatures of three cascaded devices.

Alternatively, we can replace the last two with an equivalent temperature T_{23} and then combine with the first to get:

$$T_{23} = T_2 + \frac{1}{G_2}T_3\,, \qquad T_{123} = T_1 + \frac{1}{G_1}T_{23}$$

From either point of view, we obtain the equivalent temperature:

$$\boxed{T_{123} = T_1 + \frac{1}{G_1}T_2 + \frac{1}{G_1G_2}T_3} \tag{15.8.14}$$

The system SNR will be:

$$SNR_{\text{sys}} = \frac{S_{\text{in}}}{kT_{\text{sys}}B} = \frac{S_{\text{in}}}{k(T_{\text{in}} + T_{123})B}$$

It is invariant with respect to its reference point:

$$SNR_{\text{sys}} = SNR_a = SNR_b = SNR_c = SNR_{\text{out}}$$

When expressed in terms of noise figures, Eqs. (15.8.12) and (15.8.14) are also known as Friis's formulas [1037], for example, defining the equivalent noise figure as $F_{123} = 1 + T_{123}/T_0$, we have:

$$\boxed{F_{123} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2}} \tag{15.8.15}$$

We apply now these results to the antenna receiver shown in Fig. 15.7.1, identifying the three cascaded components as the feed line, the LNA amplifier, and the rest of the receiver circuits. The corresponding noise temperatures are T_{feed}, T_{LNA}, and T_{rec}. The effective noise temperature T_{eff} of the combined system will be:

$$\boxed{T_{\text{eff}} = T_{\text{feed}} + \frac{1}{G_{\text{feed}}}T_{\text{LNA}} + \frac{1}{G_{\text{feed}}G_{\text{LNA}}}T_{\text{rec}}} \tag{15.8.16}$$

Using Eq. (15.8.8), we may replace T_{feed} in terms of the physical temperature:

$$\boxed{T_{\text{eff}} = \frac{1}{G_{\text{feed}}}(1 - G_{\text{feed}})T_{\text{phys}} + \frac{1}{G_{\text{feed}}}T_{\text{LNA}} + \frac{1}{G_{\text{feed}}G_{\text{LNA}}}T_{\text{rec}}} \tag{15.8.17}$$

The input noise temperature T_{in} to this combined system is the antenna temperature T_{ant}. It follows that *system noise temperature*, referred to either the antenna output terminals (point *a*), or to the LNA input (point *b*), will be:

$$T_{\text{sys}} = T_{sa} = T_{\text{ant}} + T_{\text{eff}} = T_{\text{ant}} + \left(\frac{1}{G_{\text{feed}}} - 1\right)T_{\text{phys}} + \frac{1}{G_{\text{feed}}}T_{\text{LNA}} + \frac{1}{G_{\text{feed}}G_{\text{LNA}}}T_{\text{rec}}$$

$$T_{sb} = G_{\text{feed}}T_{sa} = G_{\text{feed}}T_{\text{ant}} + (1 - G_{\text{feed}})T_{\text{phys}} + T_{\text{LNA}} + \frac{1}{G_{\text{LNA}}}T_{\text{rec}}$$

The importance of a high-gain low-noise amplifier is evident from Eq. (15.8.17). The high value of G_{LNA} will minimize the effect of the remaining components of the receiver system, while the small value of T_{LNA} will add only a small amount of noise. Typical values of T_{LNA} can range from 20 K for cooled amplifiers to 100 K at room temperatures.

The feed line can have a major impact. If the line is too lossy or too long, the quantity $G_{\text{feed}} = e^{-2\alpha l}$ will be small, or $1/G_{\text{feed}}$ large, contributing a significant amount to the system noise temperature. Often, the LNA is mounted *before* the feed line, near the focal point of the receiving antenna, so that the effect of the feed line will be suppressed by the factor G_{LNA}.

Similar benefits arise in base station antennas for wireless communications, where high-gain amplifiers can be placed on top of the antenna towers, instead of at the base station, which can be fairly far from the towers [1050]. Cable losses in such applications can be in the range 2–4 dB (with gain factors $G_f = 0.63$–0.40.)

The signal to system-noise ratio of the receiving system (referred to point *a* of Fig. 15.7.1) will be the ratio of the received power P_R to the system noise $N_{\text{sys}} = kT_{\text{sys}}B$. Using the Friis formula (for power transmission), we have:

$$\boxed{SNR = \frac{P_R}{N_{\text{sys}}} = \frac{P_R}{kT_{\text{sys}}B} = (P_TG_T)\frac{1}{L_f}\left(\frac{G_R}{T_{\text{sys}}}\right)\frac{1}{kB}} \tag{15.8.18}$$

This ratio is also called the *carrier-to-system-noise ratio* and is denoted by C/N. For a given transmitting EIRP, $P_T G_T$, the receiver performance depends critically on the ratio G_R/T_{sys}, referred to as the G/T ratio of the receiving antenna, or the *figure of merit*. It is usually measured in dB/K. In dB, Eq. (15.8.18) reads as:

$$\boxed{(SNR)_{\text{dB}} = (P_T G_T)_{\text{dB}} - (L_f)_{\text{dB}} + \left(\frac{G_R}{T_{\text{sys}}}\right)_{\text{dB}} - k_{\text{dB}} - B_{\text{dB}}} \tag{15.8.19}$$

The receiver SNR can be also be referred to LNA input (point *b*). The G/T ratio will not change in value, but it will be the ratio of the signal gain after the feed line divided by the system temperature T_{sb}, that is,

$$SNR = (P_T G_T)\frac{1}{L_f}\left(\frac{G_R}{T_{\text{sys}}}\right)\frac{1}{kB} = (P_T G_T)\frac{1}{L_f}\left(\frac{G_R G_{\text{feed}}}{T_{sb}}\right)\frac{1}{kB} \tag{15.8.20}$$

Example 15.8.1: Typical earth-based antennas for satellite communications have G/T ratios of the order of 40 dB/K, whereas satellite receiving antennas can have $G/T = -7$ dB/K or less. The negative sign arises from the smaller satellite antenna gain and the much higher temperature, since the satellite is looking down at a warm earth. □

Example 15.8.2: Consider a receiving antenna system as shown in Fig. 15.7.1, with antenna temperature of 40 K, feed line loss of 0.1 dB, feed line physical temperature of 290 K, LNA gain and effective noise temperature of 50 dB and 80 K. The rest of the receiver circuits have effective noise temperature of 2000 K.

Assuming the receiving antenna has a gain of 45 dB, calculate the system noise temperature and the G/T ratio at point *a* and point *b* of Fig. 15.7.1. Repeat if the feed line loss is 1 dB.

Solution: The feed line has gain $G_{\text{feed}} = 10^{-0.1/10} = 10^{-0.01} = 0.9772$, and the LNA has $G_{\text{LNA}} = 10^{50/10} = 10^5$. Thus, the system noise temperature at point *a* will be:

$$\begin{aligned} T_{\text{sys}} &= T_{\text{ant}} + \left(\frac{1}{G_{\text{feed}}} - 1\right)T_{\text{phys}} + \frac{1}{G_{\text{feed}}}T_{\text{LNA}} + \frac{1}{G_{\text{feed}}G_{\text{LNA}}}T_{\text{rec}} \\ &= 40 + \left(\frac{1}{10^{-0.01}} - 1\right)290 + \frac{80}{10^{-0.01}} + \frac{2000}{10^{-0.01}\cdot 10^5} \\ &= 40 + 6.77 + 81.87 + 0.02 = 128.62\ \text{K} = 21.09\ \text{dBK} \end{aligned}$$

At point *b*, we have $T_{sb} = G_{\text{feed}}T_{\text{sys}} = 0.9772 \times 128.62 = 125.69$ K $= 20.99$ dBK. The G/T ratio will be at point *a*, $G_R/T_{\text{sys}} = 45 - 21.09 = 23.91$ dB/K. At point *b* the gain is $G_R G_{\text{feed}} = 45 - 0.1 = 44.9$ dB, and therefore, $G/T = G_R G_{\text{feed}}/T_{sb} = 44.9 - 20.99 = 23.91$ dB/K, which is the same as at point *a*.

For a feed line loss of 1 dB, we find $T_{\text{sys}} = 215.80$ K $= 23.34$ dB. The corresponding G/T ratio will be $45 - 23.34 = 21.66$ dB. □

Example 15.8.3: Suppose the LNA were to be placed in front of the feed line of the above example. Calculate the system noise temperature in this case when the feed line loss is 0.1 dB and 1 dB.

Solution: Interchanging the roles of the feed line and the LNA in Eq. (15.8.16), we have for the system noise temperature:

$$T_{\text{sys}} = T_{\text{ant}} + T_{\text{LNA}} + \frac{1}{G_{\text{LNA}}}T_{\text{feed}} + \frac{1}{G_{\text{feed}}G_{\text{LNA}}}T_{\text{rec}}$$

With $G_{\text{feed}} = 10^{-0.1/10} = 0.9772$, we find $T_{\text{feed}} = 6.75$ K, and with $G_{\text{feed}} = 10^{-1/10} = 0.7943$, $T_{\text{feed}} = 75.1$ K. Because of the large LNA gain, the value of T_{sys} will be essentially equal to $T_{\text{ant}} + T_{\text{LNA}}$; indeed, we find in the two cases:

$$T_{\text{sys}} = 120.0205\ \text{K} \quad \text{and} \quad T_{\text{sys}} = 120.0259\ \text{K}$$

The G/T will be $45 - 10\log_{10}(120) = 20.8$ dB/K. □

15.9 Data Rate Limits

The system SNR limits the data rate between the two antennas. According to Shannon's theorem, the maximum data rate (in bits/sec) that can be achieved is:

$$C = B\log_2(1 + SNR) \qquad \text{(Shannon's channel capacity)} \tag{15.9.1}$$

where *SNR* is in absolute units. For data rates $R \le C$, Shannon's theorem states that there is an ideal coding scheme that would guarantee error-free transmission.

In a practical digital communication system, the bit-error probability or *bit-error rate* (BER), P_e, is small but not zero. The key performance parameter from which P_e can be calculated is the ratio E_b/N_0, where E_b is the energy per bit and N_0 is the system noise spectral density $N_0 = kT_{\text{sys}}$.

The functional relationship between P_e and E_b/N_0 depends on the particular digital modulation scheme used. For example, in binary and quadrature phase-shift keying (BPSK and QPSK), P_e and its inverse are given by [1047]:

$$P_e = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \quad \Leftrightarrow \quad \frac{E_b}{N_0} = \big[\text{erfinv}(1 - 2P_e)\big]^2 \tag{15.9.2}$$

where $\text{erfc}(x)$ is the complementary error function, and $\text{erf}(x)$ and $\text{erfinv}(x)$ are the error function and its inverse as defined in MATLAB:

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,dt\,, \qquad y = \text{erf}(x) \;\Leftrightarrow\; x = \text{erfinv}(y) \tag{15.9.3}$$

The relationships (15.9.2) are plotted in Fig. 15.9.1. The left graph also shows the ideal Shannon limit $E_b/N_0 = \ln 2 = 0.6931 \equiv -1.5917$ dB, which is obtained by taking the limit of Eq. (15.9.1) for infinite bandwidth.

If T_b is the time it takes to transmit one bit, then the data rate will be $R = 1/T_b$, and the required power, $P = E_b/T_b = E_b R$. It follows that the SNR will be:

$$SNR = \frac{P}{N_{\text{sys}}} = \frac{P}{kT_{\text{sys}}B} = \frac{E_b}{N_0}\frac{R}{B}$$

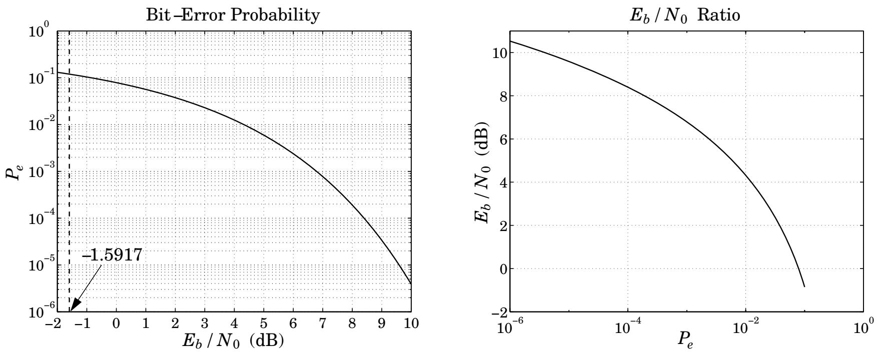


Fig. 15.9.1 P_e versus E_b/N_0, and its inverse, for a BPSK system.

Using the small-x expansion, $\log_2(1+x)\simeq x/\ln 2$, Shannon's condition for error-free transmission becomes in the limit $B\to\infty$:

$$R \le C = B\log_2\left(1+\frac{E_b}{N_0}\frac{R}{B}\right) \to B\,\frac{E_b R}{N_0 B\ln 2} = \frac{R}{\ln 2}\frac{E_b}{N_0} \quad\Rightarrow\quad \frac{E_b}{N_0} \ge \ln 2 = -1.5917\text{ dB}$$

For a pair of communicating antennas, the received power will be related to the energy per bit by $P_R = E_b/T_b = E_b R$. Using Friis's formula, we find:

$$R\,\frac{E_b}{N_0} = \frac{P_R}{N_0} = \frac{P_{\text{EIRP}}\,G_f\,G_R}{kT_{\text{sys}}} = (P_T G_T)\,\frac{G_R}{kT_{\text{sys}}}\left(\frac{\lambda}{4\pi r}\right)^2 \tag{15.9.4}$$

which may be solved for the maximum achievable data rate (in bits/sec):

$$R = \frac{1}{E_b/N_0}\,\frac{P_{\text{EIRP}}\,G_f\,G_R}{kT_{\text{sys}}} = \frac{1}{E_b/N_0}\,(P_T G_T)\,\frac{G_R}{kT_{\text{sys}}}\left(\frac{\lambda}{4\pi r}\right)^2 \tag{15.9.5}$$

An overall gain factor, $G_{\text{other}} = 1/L_{\text{other}}$, may be introduced representing other losses, such as atmospheric losses.

Example 15.9.1: The Voyager spacecrafts (launched in 1977) have antenna diameter and aperture efficiency of $d = 3.66$ m (12 ft) and $e_a = 0.6$. The operating frequency is $f = 8.415$ GHz and the transmitter power $P_T = 18$ W. Assuming the same efficiency for the 70-m receiving antenna at NASA's deep-space network at Goldstone, CA, we calculate the antenna gains using the formula $G = e_a(\pi d/\lambda)^2$, with $\lambda = c/f = 0.0357$ m:

$$G_T = 47.95\text{ dB},\quad G_R = 73.58\text{ dB},\quad P_T = 12.55\text{ dBW}$$

Assuming a system noise temperature of $T_{\text{sys}} = 25\text{ K} = 13.98$ dBK for the receiving antenna, we find for the noise spectral density $N_0 = kT_{\text{sys}} = -214.62$ dBW/Hz, where we used $k = -228.6$ dB. Assuming a bit-error rate of $P_e = 5{\times}10^{-3}$, we find from Eq. (15.9.2) the required ratio $E_b/N_0 = 3.317 = 5.208$ dB.

Voyager 1 was at Jupiter in 1979, at Saturn in 1980, and at Neptune in 1989. In 2002 it was at a distance of about $r = 12{\times}10^9$ km. It is expected to be at $r = 22{\times}10^9$ km in the year 2020. We calculate the corresponding free-space gain $G_f = (\lambda/4\pi r)^2$ and the expected data rates R from Eq. (15.9.5), where r is in units of 10^9 km:

location	r	G_f (dB)	R (dB)	R (bits/sec)
Jupiter	0.78	−288.78	49.72	93,724
Saturn	1.43	−294.05	44.45	27,885
Neptune	4.50	−304.01	34.50	2,816
2002	12.00	−312.53	25.98	396
2020	22.00	−317.79	20.71	118

where we assumed an overall loss factor of $G_{\text{other}} = -5$ dB. More information on the Voyager mission and NASA's deep-space network antennas can be obtained from the web sites [1331] and [1332]. □

15.10 Satellite Links

Consider an earth-satellite-earth system, as shown in Fig. 15.10.1. We wish to establish the total link budget and signal to system-noise ratio between the two earth antennas communicating via the satellite.

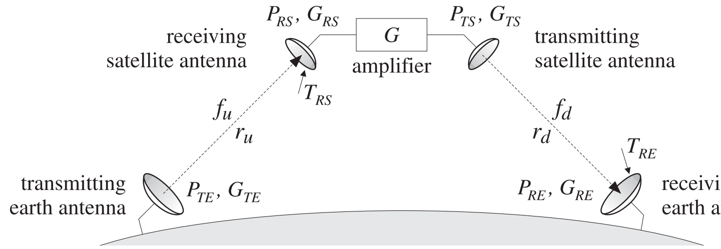


Fig. 15.10.1 Uplink and downlink in satellite communications.

In a geosynchronous satellite system, the uplink/downlink frequencies f_u, f_d are typically 6/4 GHz or 14/11 GHz. The distances r_u, r_d are of the order of 40000 km. Let $\lambda_u = c/f_u$ and $\lambda_d = c/f_d$ be the uplink/downlink wavelengths. The free-space gain/loss factors will be from Eq. (15.6.5):

$$G_{fu} = \frac{1}{L_{fu}} = \left(\frac{\lambda_u}{4\pi r_u}\right)^2,\quad G_{fd} = \frac{1}{L_{fd}} = \left(\frac{\lambda_d}{4\pi r_d}\right)^2 \tag{15.10.1}$$

The satellite has an on-board amplifier with gain G, which could be as high as 100–120 dB. Using Friis formula in its gain form, Eq. (15.6.6), the link equations for the uplink, the satellite amplification, and the downlink stages can be written as follows:

$P_{EIRP} = P_{TE}\,G_{TE}$ (EIRP of transmitting earth antenna)
$P_{RS} = P_{TE}\,G_{TE}\,G_{fu}\,G_{RS}$ (received power by satellite antenna)
$P_{TS} = GP_{RS}$ (transmitted power by satellite antenna)
$P_{RE} = P_{TS}\,G_{TS}\,G_{fd}\,G_{RE}$ (received power by earth antenna)

Expressing P_{RE} in terms of P_{TE}, we have:

$$P_{RE} = P_{RS}\,G\,G_{TS}\,G_{fd}\,G_{RE} = P_{TE}\,G_{TE}\,G_{fu}\,G_{RS}\,G\,G_{TS}\,G_{fd}\,G_{RE} \tag{15.10.2}$$

or, showing the free-space loss factors explicitly:

$$P_{RE} = P_{TE}\,G_{TE}\,G_{RS}\,G\,G_{TS}\,G_{RE}\left(\frac{\lambda_u}{4\pi r_u}\right)^2\left(\frac{\lambda_d}{4\pi r_d}\right)^2 \tag{15.10.3}$$

Because there are two receiving antennas, there will be two different system noise temperatures, say T_{RS} and T_{RE}, for the satellite and earth receiving antennas. They incorporate the antenna noise temperatures as well as the receiver components. The corresponding figures of merit will be the quantities G_{RS}/T_{RS} and G_{RE}/T_{RE}. We may define the uplink and downlink SNR's as the signal-to-system-noise ratios for the individual antennas:

$$\mathit{SNR}_u = \frac{P_{RS}}{kT_{RS}B}\,,\quad \mathit{SNR}_d = \frac{P_{RE}}{kT_{RE}B} \tag{15.10.4}$$

The system noise T_{RS} generated by the receiving satellite antenna will get amplified by G and then transmitted down to the earth antenna, where it will contribute to the system noise temperature. By the time it reaches the earth antenna it will have picked up the gain factors $G\,G_{TS}\,G_{fd}\,G_{RE}$. Thus, the net system noise temperature measured at the receiving earth antenna will be:

$$T_{\rm sys} = T_{RE} + G\,G_{TS}\,G_{fd}\,G_{RE}T_{RS} \tag{15.10.5}$$

The SNR of the total link will be therefore,

$$\mathit{SNR}_{\rm tot} = \frac{P_{RE}}{kT_{\rm sys}B} \tag{15.10.6}$$

$$\begin{aligned}\mathit{SNR}_{\rm tot}^{-1} &= \frac{k(T_{RE} + G\,G_{TS}\,G_{fd}\,G_{RE}T_{RS})B}{P_{RE}} = \frac{kT_{RE}B}{P_{RE}} + \frac{k\,G\,G_{TS}\,G_{fd}\,G_{RE}\,T_{RS}B}{P_{RE}}\\ &= \frac{kT_{RE}B}{P_{RE}} + \frac{k\,G\,G_{TS}\,G_{fd}\,G_{RE}T_{RS}B}{G\,G_{TS}\,G_{fd}\,G_{RE}P_{RS}} = \frac{kT_{RE}B}{P_{RE}} + \frac{kT_{RS}B}{P_{RS}} = \mathit{SNR}_d^{-1} + \mathit{SNR}_u^{-1}\end{aligned}$$

where we used Eq. (15.10.2). It follows that:

$$\boxed{\mathit{SNR}_{\rm tot} = \frac{1}{\mathit{SNR}_u^{-1} + \mathit{SNR}_d^{-1}}} \tag{15.10.7}$$

This is also written in the form:

$$\left(\frac{C}{N}\right)_{\rm tot} = \frac{1}{\left(\dfrac{C}{N}\right)_u^{-1} + \left(\dfrac{C}{N}\right)_d^{-1}}$$

Example 15.10.1: As an example of a link budget calculation, assume the following data: The uplink/downlink distances are 36000 km. The uplink/downlink frequencies are 6/4 GHz. The diameters of the earth and satellite antennas are 15 m and 0.5 m with 60% aperture efficiencies. The earth antenna transmits power of 1 kW and the satellite transponder gain is 90 dB. The satellite receiving antenna is looking down at an earth temperature of 300 K and has a noisy receiver of effective noise temperature of 2700 K, whereas the earth receiving antenna is looking up at a sky temperature of 50 K and uses a high-gain LNA amplifier of 80 K (feedline losses may be ignored.) The bandwidth is 30 MHz.

The uplink and downlink wavelengths are $\lambda_u = 0.05$ m and $\lambda_d = 0.075$ m, corresponding to 6 and 4 GHz. The up and down free-space gains and losses are:

$$G_{fu} = -L_{fu} = -199.13 \text{ dB},\quad G_{fd} = -L_{fd} = -195.61 \text{ dB}$$

The antenna gains are calculated to be:

$$G_{TE} = 57.27 \text{ dB},\quad G_{RS} = 27.72 \text{ dB},\quad G_{TS} = 24.20 \text{ dB},\quad G_{RE} = 53.75 \text{ dB}$$

With $P_{TE} = 1$ kW $= 30$ dBW, the EIRP of the transmitting earth antenna will be: $P_{\rm EIRP} = 30 + 57.27 = 87.27$ dBW. The power received by the satellite antenna will be:

$$P_{RS} = 87.27 - 199.13 + 27.72 = -84.14 \text{ dBW}$$

After boosting this up by the transponder gain of 90 dB, the power transmitted down to the receiving earth antenna will be:

$$P_{TS} = 90 - 84.14 = 5.86 \text{ dBW}$$

The EIRP of the transmitting satellite antenna will be $(P_{TS}G_{TS})_{\rm dB} = 5.86 + 24.20 = 30.06$ dBW. The downlink power received by the earth antenna will be:

$$P_{RE} = 30.06 - 195.61 + 53.75 = -111.80 \text{ dBW}$$

The system noise temperatures are: $T_{RS} = 300 + 2700 = 3000$ K and $T_{RE} = 50 + 80 = 130$ K, and in dBK: $T_{RS} = 34.77$ and $T_{RE} = 21.14$. The 30 MHz bandwidth is in dB: $B_{\rm dB} = 10\log_{10}(30\times 10^6) = 74.77$ dB Hz. Using the Boltzmann constant k in dB, $k_{\rm dB} = -228.6$, we calculate the receiver system noise powers in dB, using $N = k_{\rm dB} + T_{\rm dB} + B_{\rm dB}$:

$$\begin{aligned}N_{RS} &= -228.6 + 34.77 + 74.77 = -119.06 \text{ dBW}\\ N_{RS} &= -228.6 + 21.14 + 74.77 = -132.69 \text{ dBW}\end{aligned}$$

It follows that the G/T ratios and system SNR's for the receiving antennas will be:

$$(G/T)_u = G_{RS} - T_{RS} = 27.72 - 34.77 = -7.05 \text{ dB/K}$$

$$(G/T)_d = G_{RE} - T_{RE} = 53.75 - 21.14 = 32.61 \text{ dB/K}$$

$$\mathit{SNR}_u = P_{RS} - N_{RS} = -84.14 + 119.06 = 34.92 \text{ dB} = 3103.44$$

$$\mathit{SNR}_d = P_{RE} - N_{RE} = -111.80 + 132.69 = 20.89 \text{ dB} = 122.72$$

The overall system SNR is calculated from Eq. (15.10.7) using absolute units:

$$\mathit{SNR}_{\rm tot} = \frac{1}{\mathit{SNR}_u^{-1} + \mathit{SNR}_d^{-1}} = \frac{1}{(3103.44)^{-1} + (122.72)^{-1}} = 118.05 = 20.72 \text{ dB}$$

The overall SNR is essentially equal to the downlink SNR. □

15.11 Radar Equation

Another example of the application of the concepts of gain and effective area and the use of Friis formulas is radar. Fig. 15.11.1 shows a radar antenna, which illuminates a target at distance r in the direction of its maximal gain. The incident wave on the target will be reflected and a portion of it will be intercepted back at the antenna.

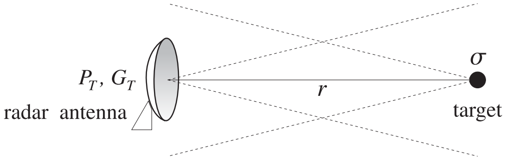


Fig. 15.11.1 Radar antenna and target.

The concept of *radar cross section* σ provides a measure of the effective area of the target and the re-radiated power. If the radar antenna transmits power P_T with gain G_T, the power density of the transmitted field at the location of the target will be:

$$\boxed{\mathcal{P}_T = \frac{P_T G_T}{4\pi r^2}} \tag{15.11.1}$$

From the definition of σ, the power intercepted by the target and re-radiated is:

$$\boxed{P_{\text{target}} = \sigma \mathcal{P}_T = \frac{P_T G_T \sigma}{4\pi r^2}} \tag{15.11.2}$$

By definition of the radar cross section, the power P_{target} will be re-radiated isotropically and establish a power density back at the location of the radar antenna:

$$\boxed{\mathcal{P}_{\text{target}} = \frac{P_{\text{target}}}{4\pi r^2} = \frac{P_T G_T \sigma}{(4\pi r^2)^2}} \tag{15.11.3}$$

The amount of power received by the radar antenna will be given in terms of its effective area A_R as follows:

$$\boxed{P_R = A_R \mathcal{P}_{\text{target}} = \frac{P_T G_T A_R \sigma}{(4\pi)^2 r^4}} \qquad \text{(radar equation)} \tag{15.11.4}$$

This is also known as Friis' formula. Using $A_R = A_T$ and $G_T = 4\pi A_T/\lambda^2$, we may express Eq. (15.11.4) in the alternative forms:

$$P_R = \frac{P_T A_T^2 \sigma}{4\pi \lambda^2 r^4} = \frac{P_T G_T^2 \lambda^2 \sigma}{(4\pi)^3 r^4} = P_T G_T^2 \left(\frac{\lambda}{4\pi r}\right)^4 \left(\frac{4\pi\sigma}{\lambda^2}\right) \tag{15.11.5}$$

Introducing the equivalent target gain corresponding to the radar cross section, that is, $G_\sigma = 4\pi\sigma/\lambda^2$, we may also write Eq. (15.11.5) as the product of gains:

$$P_R = P_T G_T^2 G_f^2 G_\sigma \tag{15.11.6}$$

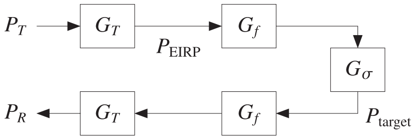


Fig. 15.11.2 Gain model of radar equation.

Fig. 15.11.2 shows this gain model. There are two free-space paths and two antenna gains, acting as transmit and receive gains.

The minimum detectable received power, $P_{R,\min}$, defines the maximum distance $r_{\max}$ at which the target can be detected:

$$P_{R,\min} = \frac{P_T G_T A_R \sigma}{(4\pi)^2 r_{\max}^4}$$

Solving for $r_{\max}$, we obtain:

$$\boxed{r_{\max} = \left[\frac{P_T G_T A_T \sigma}{(4\pi)^2 P_{R,\min}}\right]^{1/4}} \qquad \text{(radar range)} \tag{15.11.7}$$

If the target is not in the direction of maximal gain G_T of the antenna, but in some other direction, say (θ,ϕ), then the maximal gain G_T in Eq. (15.11.5) must be replaced with $G_T g(\theta,\phi)$, where $g(\theta,\phi)$ is the antenna's normalized gain. The received power can be expressed then as:

$$P_R = \frac{P_T G_T^2 g^2(\theta,\phi) \lambda^2 \sigma}{(4\pi)^3 r^4} \tag{15.11.8}$$

In ground-based air search radars trying to detect approaching aircraft flying at a fixed height h, the power received by the radar can be made to be *independent* of the distance r, within a certain distance range, by choosing the gain $g(\theta,\phi)$ appropriately. As shown in Fig. 15.11.3, the height h is related to r by $h = r\cos\theta$.

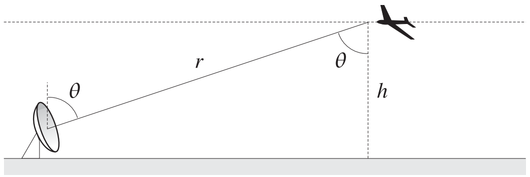


Fig. 15.11.3 Secant antenna gain.

If the gain is designed to have the *secant-squared* shape $g(\theta,\phi) = K/\cos^2\theta$, where K is a constant, then the power will become independent of r. Indeed,

$$P_R = \frac{P_T G_T^2 g^2(\theta,\phi) \lambda^2 \sigma}{(4\pi)^3 r^4} = \frac{P_T G_T^2 K^2 \lambda^2 \sigma}{(4\pi)^3 r^4 \cos^4\theta} = \frac{P_T G_T^2 K^2 \lambda^2 \sigma}{(4\pi)^3 h^4}$$

The secant behavior is not valid over all polar angles θ, but only over a certain range, such as $0 \le \theta \le \theta_{\max}$, where $\theta_{\max}$ corresponds to the maximum range of the radar $r_{\max} = h/\cos\theta_{\max}$. The desired secant shape can be achieved by appropriate feeds of the radar dish antenna, or by an antenna array with properly designed array factor. In Sec. 20.5, we present such a design for an array.

15.12 Problems

15.1 In an earth-satellite-earth communication system, the uplink/downlink distances are 36000 km. The uplink/downlink frequencies are 6/4 GHz. The diameters of the earth and satellite antennas are 20 m and 1 m with 60% aperture efficiencies. The transmitting earth antenna transmits power of 1.5 kW. The satellite transponder gain is 85 dB. The satellite receiving antenna is looking down at an earth temperature of 290K and has a noisy receiver of effective noise temperature of 3000K, whereas the earth receiving antenna is looking up at a sky temperature of 60K and uses a high-gain LNA amplifier of noise temperature of 100K (feedline losses may be ignored.) The bandwidth is 30 MHz.

a. Calculate all antenna gains in dB.

b. Calculate the uplink and downlink free-space losses in dB.

c. Calculate the amount of power received by the satellite in dBW. Calculate the uplink signal to noise ratio in dB.

d. Calculate the power received by the receiving earth antenna in dBW and the downlink signal to noise ratio.

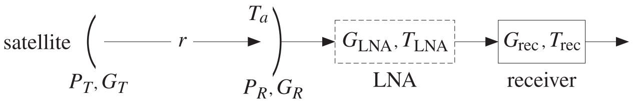
e. Finally, calculate the total system signal to noise ratio in dB.

15.2 The Voyager spacecraft is currently transmitting data to earth from a distance of 12 billion km. Its antenna diameter and aperture efficiency are 3.66 m and 60%. The operating frequency is 8.415 GHz and Voyager's transmitter power is 18 W. Assume the same aperture efficiency for the 70-m receiving antenna at NASA's deep-space network at Goldstone, CA.

a. Calculate the spacecraft's and earth's antenna gains in dB. Calculate also the free-space loss in dB.

b. Calculate the achievable communication data rate in bits/sec between Voyager and earth using QPSK modulation and assuming the following: an overall transmission loss factor of 5 dB, a system noise temperature of 25 K, an energy-per-bit to noise-spectral-density ratio of $E_b/N_0 = 3.317 = 5.208$ dB, which for QPSK corresponds to a bit-error probability of $P_e = 5{\times}10^{-3}$.

15.3 A satellite to earth downlink (shown below) is operating at a carrier frequency of f Hertz using QPSK modulation and achieving a bit rate of R bits/sec with a bit error probability of P_e. With the LNA absent, the receiving earth antenna is connected directly to a noisy receiver with equivalent noise temperature of T_{rec}. Both antennas are dishes.


satellite
r
T_a
$G_{\text{LNA}}, T_{\text{LNA}}$
$G_{\text{rec}}, T_{\text{rec}}$
P_T, G_T
P_R, G_R
LNA
receiver

1. A low-noise amplifier of very high gain G_{LNA} and low noise temperature T_{LNA} is inserted between the earth antenna and the receiver. Show that the presence of the LNA allows the link to be operated (with the same error probability P_e) at the higher bit rate:

$$R_{\text{new}} = R\,\frac{T_a + T_{\text{rec}}}{T_a + T_{\text{LNA}}}$$

where T_a is the earth antenna noise temperature, and $T_{\text{LNA}} \ll T_{\text{rec}}$.

2. The equation in part (a) is an approximation. Derive the exact form of that equation and discuss the nature of the approximation that was made.

3. How would the expression in part (a) change if, in addition to the assumptions of part (a), the operating frequency f were to be doubled? Explain your reasoning. How would (a) change if the transmitter power P_T were to double? If the distance r were to double?

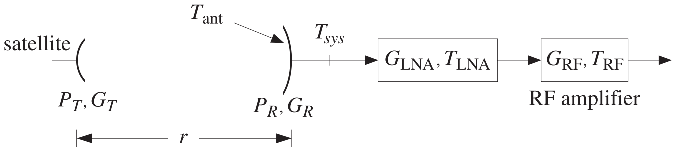
4. With the LNA present, and assuming that the bit rate R, error probability P_e, and f, P_T, r remain the same, show that the diameter d of the earth antenna can be lowered to the following value without affecting performance:

$$d_{\text{new}} = d\,\sqrt{\frac{T_a + T_{\text{LNA}}}{T_a + T_{\text{rec}}}}$$

where the same approximation was made as in part (a).

15.4 A satellite to earth link (shown below) is operating at the carrier frequency of 4 GHz. The data link employs QPSK modulation and achieves a bit-error-rate probability of $P_e = 10^{-6}$. The satellite has transmitter power of 20 W and uses a dish antenna that has a diameter of 0.5 m and aperture efficiency of 0.6. The earth antenna has diameter of 2 m, efficiency of 0.6, and antenna noise temperature of 80 K. The satellite antenna is at a distance of 40,000 km from the earth antenna.

The output of the receiving antenna is connected to a high-gain low noise amplifier with gain of 40 dB and equivalent noise temperature of 200 K. The output of the LNA is connected to an RF amplifier with equivalent noise temperature of 1800 K.



For QPSK modulation, we have the relationship $P_e = \text{erfc}\big(\sqrt{E_b/N_0}\big)/2$ with inverse $E_b/N_0 = [\text{erfinv}(1-2P_e)]^2$. For the purposes of this exam, the following equation provides an excellent approximation to this inverse relationship over the range of $10^{-8} \le P_e \le 10^{-3}$:

$$\frac{E_b}{N_0} = -2.1969\,\log_{10}(P_e) - 1.8621$$

where E_b/N_0 is in absolute units.

a. Calculate the achievable communication data rate R in megabits/sec.

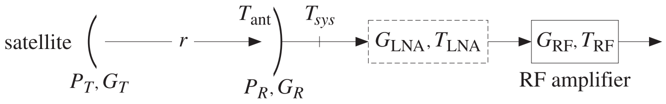
b. If the LNA is removed, the performance of the system will deteriorate. In an attempt to keep the data rate the same as in part (a), the satellite transmitter power is increased to 80 W. Calculate the deteriorated value of the bit-error-rate P_e in this case.

15.5 A satellite to earth downlink (shown below) is operating at the carrier frequency of 4 GHz. The distance between the two antennas is $r = 40\,000$ km. The bit error probability is $P_e = 10^{-5}$ using QPSK modulation.

For QPSK modulation, we have the following relationship between the bit-error-probability and E_b/N_0 ratio, expressed in terms of the MATLAB functions `erfc` and `erfinv`:

$$P_e = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \quad \Leftrightarrow \quad \frac{E_b}{N_0} = \big[\text{erfinv}(1-2P_e)\big]^2$$

The satellite has transmitter power of 20 W and uses a dish antenna that has a diameter of 0.5 m and aperture efficiency of 0.6. The earth antenna has diameter of 5 m, efficiency of 0.6, and antenna noise temperature of 50 K. The output of the antenna is connected to an RF amplifier with equivalent noise temperature of 2000 K.



a. Assuming that no LNA is used, calculate the system noise temperature T_{sys} at the output of the receiving antenna, the received power P_R in picowatts, and the maximum achievable data rate in Mb/sec.

b. It is desired to improved the performance of this system tenfold, that is, to increase the maximum achievable data rate in Mb/sec by a factor of 10. To this end, a low-noise amplifier of 40-dB gain is inserted as shown. Determine the noise temperature of the LNA that would guarantee such a performance improvement.

c. What is the maximum noise temperature of the LNA that can achieve such a 10-fold improvement, and at what LNA gain is it achieved?

15.6 A radar with EIRP of $P_{\text{radar}} = P_T G_T$ is trying to detect an aircraft of radar cross section σ. The aircraft is at a distance r from the radar and tries to conceal itself by jamming the radar with an on-board jamming antenna of EIRP of $P_{\text{jammer}} = P_J G_J$. Assume that both the radar and the jamming antennas are pointing in their direction of maximal gains.

a. Derive an expression of the signal-to-jammer ratio S/J, where S represents the power received from the target back at the radar antenna according to the radar equation, and J represents the power from the jamming antenna received by the radar antenna. Express the ratio in terms of P_{radar}, P_{jammer}, r, and σ.

b. If detectability of the target in the presence of jamming requires at least a 0-dB signal-to-jammer ratio (that is, $S/J \geq 1$), show that the maximum detectable distance is:

$$r = \sqrt{\frac{P_T G_T}{P_J G_J}\,\frac{\sigma}{4\pi}}$$

15.7 The Arecibo Observatory in Puerto Rico has a gigantic dish antenna of diameter of 1000 ft (304.8 m). It transmits power of 2.5 MW at a frequency of 430 MHz.

a. Assuming a 60 percent effective area, what is its gain in dB?

b. What is its beamwidth in degrees?

c. If used as a radar and the minimum detectable received power is −130 dBW, what is its maximum range for detecting a target of radar cross-section of 1 m^2?

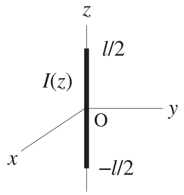
16

Linear and Loop Antennas

16.1 Linear Antennas

The radiation angular pattern of antennas is completely determined by the transverse component $\boldsymbol{F}_\perp = \hat{\boldsymbol{\theta}} F_\theta + \hat{\boldsymbol{\phi}} F_\phi$ of the radiation vector $\boldsymbol{F}$, which in turn is determined by the current density $\boldsymbol{J}$. Here, we consider some examples of current densities describing various antenna types, such as linear antennas, loop antennas, and linear arrays.

For *linear antennas*, we may choose the z-axis to be along the direction of the antenna. Assuming an infinitely thin antenna, the current density will have the form:



$$\boxed{\boldsymbol{J}(\mathbf{r}) = \hat{\mathbf{z}}\, I(z)\,\delta(x)\,\delta(y)} \qquad \text{(thin linear antenna)} \tag{16.1.1}$$

where $I(z)$ is the current distribution along the antenna element. It is shown in Sec. 21.4 that $I(z)$ satisfies approximately the Helmholtz equation along the antenna:

$$\frac{d^2 I(z)}{dz^2} + k^2 I(z) = 0 \tag{16.1.2}$$

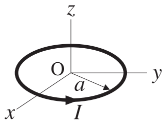
Some examples of current distributions $I(z)$ are as follows:

$I(z) = Il\delta(z)$	Hertzian dipole
$I(z) = I$	Uniform line element
$I(z) = I(1 - 2\lvert z\rvert/l)$	Small linear dipole
$I(z) = I\sin\big(k(l/2 - \lvert z\rvert)\big)$	Standing-wave antenna
$I(z) = I\cos(kz)$	Half-wave antenna $(l = \lambda/2)$
$I(z) = Ie^{-jkz}$	Traveling-wave antenna

where l is the length of the antenna element and the expressions are assumed to be valid for $-l/2 \le z \le l/2$, so that the antenna element straddles the xy-plane.

The Hertzian dipole, uniform line element, and small linear dipole examples do not satisfy Eq. (16.1.2), except when the antenna length is electrically *short*, that is, $l \ll \lambda$.

For *loop antennas*, we may take the loop to lie on the xy-plane and be centered at the origin. Again, we may assume a thin wire. For a circular loop of radius a, the current flows azimuthally. The corresponding current density can be expressed in cylindrical coordinates $\mathbf{r} = (\rho, \phi, z)$ as:



$$\boxed{\boldsymbol{J}(\mathbf{r}) = \hat{\boldsymbol{\phi}}\, I\delta(\rho - a)\,\delta(z)} \qquad \text{(circular loop)} \tag{16.1.3}$$

The delta functions confine the current on the $\rho = a$ circle on the xy-plane. We will discuss loop antennas in Sec. 16.8.

Antenna arrays may be formed by considering a group of antenna elements, such as Hertzian or half-wave dipoles, arranged in particular geometrical configurations, such as along a particular direction. Some examples of antenna arrays that are made up from identical antenna elements are as follows:

$$
\begin{aligned}
\boldsymbol{J}(\mathbf{r}) &= \hat{\mathbf{z}} \sum_n a_n I(z)\,\delta(x - x_n)\,\delta(y) && \text{array along } x\text{-direction} \\
\boldsymbol{J}(\mathbf{r}) &= \hat{\mathbf{z}} \sum_n a_n I(z)\,\delta(y - y_n)\,\delta(x) && \text{array along } y\text{-direction} \\
\boldsymbol{J}(\mathbf{r}) &= \hat{\mathbf{z}} \sum_n a_n I(z - z_n)\,\delta(x)\,\delta(y) && \text{array along } z\text{-direction} \\
\boldsymbol{J}(\mathbf{r}) &= \hat{\mathbf{z}} \sum_{mn} a_{mn} I(z)\,\delta(x - x_m)\,\delta(y - y_n) && \text{2D planar array}
\end{aligned}
$$

The weights a_n, a_{mn} are chosen appropriately to achieve desired directivity properties for the array. We discuss arrays in Sec. 19.1.

It is evident now from Eq. (16.1.1) that the radiation vector $\boldsymbol{F}$ will have only a z-component. Indeed, we have from the definition Eq. (14.7.5):

$$\boldsymbol{F} = \int_V \boldsymbol{J}(\mathbf{r}')\, e^{j\,\mathbf{k}\cdot\mathbf{r}'}\, d^3\mathbf{r}' = \hat{\mathbf{z}} \int I(z')\,\delta(x')\,\delta(y')\, e^{j(k_x x' + k_y y' + k_z z')}\, dx'\, dy'\, dz'$$

The x' and y' integrations are done trivially, whereas the z' integration extends over the length l of the antenna. Thus,

$$\boldsymbol{F} = \hat{\mathbf{z}}\, F_z = \hat{\mathbf{z}} \int_{-l/2}^{l/2} I(z')\, e^{jk_z z'}\, dz'$$

Using Eq. (14.8.3), the wave vector $\boldsymbol{k}$ can be resolved in cartesian components as:

$$\boldsymbol{k} = k\,\hat{\mathbf{r}} = \hat{\mathbf{x}}\, k\cos\phi\sin\theta + \hat{\mathbf{y}}\, k\sin\phi\sin\theta + \hat{\mathbf{z}} k\cos\theta = \hat{\mathbf{x}}\, k_x + \hat{\mathbf{y}}\, k_y + \hat{\mathbf{z}}\, k_z$$

Thus,

$$\boxed{\begin{aligned} k_x &= k\cos\phi\sin\theta \\ k_y &= k\sin\phi\sin\theta \\ k_z &= k\cos\theta \end{aligned}} \tag{16.1.4}$$

It follows that the radiation vector F_z will only depend on the polar angle θ:

$$\boxed{F_z(\theta) = \int_{-l/2}^{l/2} I(z')\, e^{jk_z z'}\, dz' = \int_{-l/2}^{l/2} I(z')\, e^{jkz'\cos\theta}\, dz'} \tag{16.1.5}$$

Using Eq. (14.8.2) we may resolve $\hat{\mathbf{z}}$ into its spherical coordinates and identify the radial and transverse components of the radiation vector:

$$\boldsymbol{F} = \hat{\mathbf{z}}\, F_z = (\hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta)\, F_z(\theta) = \hat{\mathbf{r}}\, F_z(\theta)\cos\theta - \hat{\boldsymbol{\theta}}\, F_z(\theta)\sin\theta$$

Thus, the transverse component of $\boldsymbol{F}$ will be have only a θ-component:

$$\boldsymbol{F}_\perp(\theta) = \hat{\boldsymbol{\theta}}\, F_\theta(\theta) = -\hat{\boldsymbol{\theta}}\, F_z(\theta)\sin\theta$$

It follows that the electric and magnetic radiation fields (14.10.5) generated by a linear antenna will have the form:

$$\boxed{\begin{aligned} \boldsymbol{E} &= \hat{\boldsymbol{\theta}}\, E_\theta = \hat{\boldsymbol{\theta}}\, jk\eta\, \frac{e^{-jkr}}{4\pi r}\, F_z(\theta)\sin\theta \\ \boldsymbol{H} &= \hat{\boldsymbol{\phi}}\, H_\phi = \hat{\boldsymbol{\phi}}\, jk\, \frac{e^{-jkr}}{4\pi r}\, F_z(\theta)\sin\theta \end{aligned}} \tag{16.1.6}$$

The fields are *omnidirectional*, that is, independent of the azimuthal angle ϕ. The factor $\sin\theta$ arises from the cartesian to spherical coordinate transformation, whereas the factor $F_z(\theta)$ incorporates the dependence on the assumed current distribution $I(z)$. The radiation intensity $U(\theta, \phi)$ has θ-dependence only and is given by Eq. (15.1.4):

$$\boxed{U(\theta) = \frac{\eta k^2}{32\pi^2}\, |F_z(\theta)|^2 \sin^2\theta} \qquad \text{(radiation intensity of linear antenna)} \tag{16.1.7}$$

To summarize, the radiated fields, the total radiated power, and the angular distribution of radiation from a linear antenna are completely determined by the quantity $F_z(\theta)$ defined in Eq. (16.1.5).

16.2 Hertzian Dipole

The simplest linear antenna example is the Hertzian dipole that has a current distribution $I(z) = Il\delta(z)$ corresponding to an infinitesimally small antenna located at the origin. Eq. (16.1.5) yields:

$$F_z(\theta) = \int_{-l/2}^{l/2} I(z')\, e^{jk_z z'}\, dz' = \int_{-l/2}^{l/2} Il\delta(z')\, e^{jkz'\cos\theta}\, dz' = Il$$

Thus, F_z is a constant independent of θ. The radiation intensity is obtained from Eq. (16.1.7):

$$U(\theta) = \frac{\eta k^2}{32\pi^2}\, |Il|^2 \sin^2\theta$$

Its maximum occurs at $\theta = \pi/2$, that is, broadside to the antenna:

$$U_{\max} = \frac{\eta k^2}{32\pi^2}|Il|^2$$

It follows that the normalized power gain will be:

$$\boxed{g(\theta) = \frac{U(\theta)}{U_{\max}} = \sin^2\theta} \qquad \text{(Hertzian dipole gain)} \tag{16.2.1}$$

The gain $g(\theta)$ is plotted in absolute and dB units in Fig. 16.2.1. Note that the 3-dB or half-power circle intersects the gain curve at 45° angles. Therefore, the *half-power beam width* (HPBW) will be 90°—not a very narrow beam. We note also that there is no radiated power *along* the direction of the antenna element, that is, the z-direction, or $\theta = 0$.

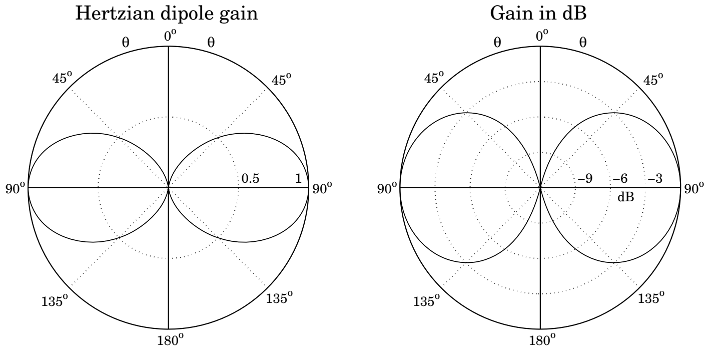


Fig. 16.2.1 Gain of Hertzian dipole in absolute and dB units.

In these plots, the gain was computed by the function `dipole` and plotted with `abp` and `dbp`. For example the left figure was generated by:

```
[g, th, c] = dipole(0, 200);
abp(th, g, 45);
```

Next, we calculate the beam solid angle from:

$$\Delta\Omega = \int_0^{\pi}\int_0^{2\pi} g(\theta)\sin\theta\, d\theta\, d\phi = 2\pi\int_0^{\pi} g(\theta)\sin\theta\, d\theta = 2\pi\int_0^{\pi}\sin^3\theta\, d\theta\,, \qquad \text{or,}$$

$$\Delta\Omega = \frac{8\pi}{3}$$

It follows that the directivity will be:

$$D_{\max} = \frac{4\pi}{\Delta\Omega} = \frac{4\pi}{8\pi/3} = 1.5 \equiv 1.76\text{ dB}$$

The total radiated power is then found from Eq. (15.2.17):

$$P_{\text{rad}} = U_{\max}\Delta\Omega = \frac{\eta k^2}{32\pi^2}|Il|^2\frac{8\pi}{3} = \frac{\eta k^2|Il|^2}{12\pi} \tag{16.2.2}$$

Because of the proportionality to $|I|^2$, we are led to define the *radiation resistance* of the antenna, R_{rad}, as the resistance that would dissipate the same amount of power as the power radiated, that is, we define it through:

$$\boxed{P_{\text{rad}} = \frac{1}{2}R_{\text{rad}}|I|^2} \tag{16.2.3}$$

Comparing the two expressions for P_{rad}, we find:

$$R_{\text{rad}} = \frac{\eta k^2 l^2}{6\pi} = \frac{2\pi\eta}{3}\left(\frac{l}{\lambda}\right)^2 \tag{16.2.4}$$

where we replaced $k = 2\pi/\lambda$. Because we assumed an infinitesimally small antenna, $l \ll \lambda$, the radiation resistance will be very small.

A related antenna example is the finite Hertzian, or uniform line element, which has a constant current I flowing along its entire length l, that is, $I(z) = I$, for $-l/2 \le z \le l/2$. We can write $I(z)$ more formally with the help of the unit-step function $u(z)$ as follows:

$$I(z) = I\,[u(z + l/2) - u(z - l/2)]$$

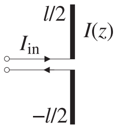
The Hertzian dipole may be thought of as the limiting case of this example in the limit $l \to 0$. Indeed, multiplying and dividing by l, and using the property that the derivative of the unit-step is $u'(z) = \delta(z)$, we have

$$I(z) = Il\,\frac{u(z + l/2) - u(z - l/2)}{l} \to Il\frac{du(z)}{dz} = Il\delta(z)$$

and we must assume, of course, that the product Il remains finite in that limit.

16.3 Standing-Wave Antennas

A very practical antenna is the center-fed standing-wave antenna, and in particular, the half-wave dipole whose length is $l = \lambda/2$. The current distribution along the antenna length is assumed to be a standing wave, much like the case of an open-ended parallel wire transmission line. Indeed, as suggested by the figure below, the center-fed dipole may be thought of as an open-ended transmission line whose ends have been bent up and down. The current distribution is:



$$I(z) = I\sin\big(k(l/2 - |z|)\big) \qquad \text{(standing-wave antenna)} \tag{16.3.1}$$

Defining the half-length $h = l/2$, the radiation vector z-component $F_z(\theta)$ is:

$$F_z(\theta)=\int_{-h}^{h} I\sin\big(k(l/2-|z'|)\big)e^{jkz'\cos\theta}dz' = \frac{2I}{k}\,\frac{\cos(kh\cos\theta)-\cos(kh)}{\sin^2\theta}$$

Inserting $F_z(\theta)$ into Eq. (16.1.7), and canceling some common factors, we obtain:

$$U(\theta)=\frac{\eta|I|^2}{8\pi^2}\left|\frac{\cos(kh\cos\theta)-\cos(kh)}{\sin\theta}\right|^2 \tag{16.3.2}$$

It follows that the normalized power gain $g(\theta)$ will have a similar form:

$$g(\theta)=c_n\left|\frac{\cos(kh\cos\theta)-\cos(kh)}{\sin\theta}\right|^2 \qquad \text{(normalized gain)} \tag{16.3.3}$$

where c_n is a normalization constant chosen to make the maximum of $g(\theta)$ equal to unity. Depending on the value of l, this maximum may not occur at $\theta = \pi/2$.

In the limit $l \to 0$, we obtain the gain of the Hertzian dipole, $g(\theta)=\sin^2\theta$. For small values of l, we obtain the linear-current case. Indeed, using the approximation $\sin x \simeq x$, the current (16.3.1) becomes:

$$I(z)=Ik\left(\frac{l}{2}-|z|\right), \qquad -\frac{l}{2}\le z\le\frac{l}{2}$$

For a general dipole of length l, the current at the input terminals of the antenna is not necessarily equal to the peak amplitude I. Indeed, setting $z = 0$ in (16.3.1) we have:

$$I_{\text{in}} = I(0)=I\sin(kl/2)=I\sin kh \tag{16.3.4}$$

The radiation resistance may be defined either in terms of the peak current or in terms of the input current through the definitions:

$$P_{\text{rad}} = \frac{1}{2}R_{\text{peak}}|I|^2 = \frac{1}{2}R_{\text{in}}|I_{\text{in}}|^2 \quad\Rightarrow\quad R_{\text{in}} = \frac{R_{\text{peak}}}{\sin^2 kh} \tag{16.3.5}$$

When l is a half-multiple of λ, the input and peak currents are equal and the two definitions of the radiation resistance are the same. But when l is a multiple of λ, Eq. (16.3.4) gives zero for the input current, which would imply an infinite input resistance R_{in}. In practice, the current distribution is only approximately sinusoidal and the input current is not exactly zero.

The input impedance of an antenna has in general both a resistive part R_{in} and a reactive part X_{in}, so that $Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}}$. The relevant theory is discussed in Sec. 22.2. Assuming a sinusoidal current, Z_{in} can be computed by Eq. (22.2.10), implemented by the MATLAB function `imped`:

```
Zin = imped(l,a);          % input impedance of standing-wave antenna
```

where l, a are the length and radius of the antenna in units of λ. For example, a half-wave dipole ($l = \lambda/2$) with zero radius has $Z_{\text{in}} = \text{imped}(0.5, 0)= 73.1 + j\,42.5\ \Omega$.

For $l \gg a$, the input resistance remains largely independent of the radius a. The reactance has a stronger dependence on a. Fig. 16.3.1 shows a plot of R_{in} and X_{in} versus the antenna length l plotted over the interval $0.3\lambda \le l \le 0.7\lambda$, for the three choices of the radius: $a = 0$, $a = 0.0005\lambda$, and $a = 0.005\lambda$.

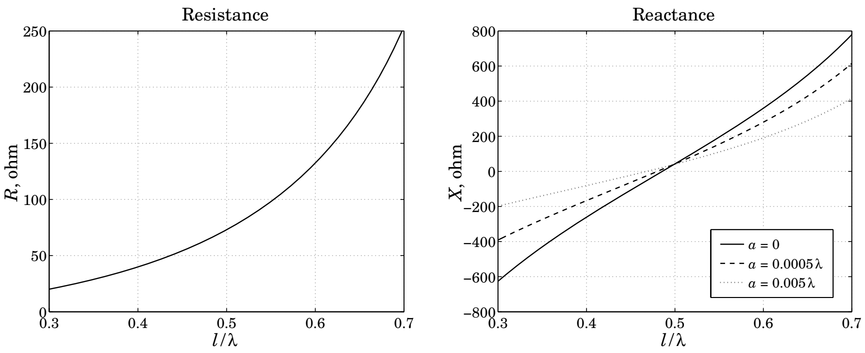


Fig. 16.3.1 Input impedance of standing-wave dipole antenna.

We observe that the reactance X_{in} vanishes for lengths that are a little shorter than $l = \lambda/2$. Such antennas are called *resonant antennas* in analogy with a resonant RLC circuit whose input impedance $Z = R + j(\omega L - 1/\omega C)$ has a vanishing reactance at its resonant frequency $\omega = 1/\sqrt{LC}$.

For the three choices of the radius a, we find the following resonant lengths and corresponding input resistances:

$$\begin{array}{lll} a = 0, & l = 0.4857\lambda, & R_{\text{in}} = 67.2\ \Omega \\ a = 0.0005\lambda, & l = 0.4801\lambda, & R_{\text{in}} = 65.0\ \Omega \\ a = 0.005\lambda, & l = 0.4681\lambda, & R_{\text{in}} = 60.5\ \Omega \end{array}$$

An analytical expression for the peak and input radiation resistances can be obtained by integrating the radiation intensity (16.3.2) over all solid angles to get the total radiated power:

$$\begin{aligned} P_{\text{rad}} &= \int U(\theta)\,d\Omega = \int_0^{\pi}\int_0^{2\pi} U(\theta)\sin\theta\,d\theta\,d\phi = 2\pi\int_0^{\pi} U(\theta)\sin\theta\,d\theta \\ &= \frac{\eta|I|^2}{4\pi}\int_0^{\pi}\frac{\big(\cos(kh\cos\theta)-\cos(kh)\big)^2}{\sin\theta}\,d\theta \end{aligned}$$

Comparing with (16.3.5), we obtain the peak resistance:

$$R_{\text{peak}} = \frac{\eta}{2\pi}\int_0^{\pi}\frac{\big(\cos(kh\cos\theta)-\cos(kh)\big)^2}{\sin\theta}\,d\theta$$

Using the trigonometric identity,

$$\begin{aligned} &\big(\cos(kh\cos\theta)-\cos(kh)\big)^2 \\ &\quad= \frac{1}{2}\big(\cos(2kh\cos\theta)-\cos(2kh)\big) - 2\big(\cos(kh\cos\theta)-\cos(kh)\big)\cos kh \end{aligned}$$

the above integral can be expressed as a sum of two integrals of the form:

$$\int_0^\pi \frac{\cos(\alpha\cos\theta)-\cos\alpha}{\sin\theta}\,d\theta = S_i(2\alpha)\sin\alpha - C_{\rm in}(2\alpha)\cos\alpha$$

which is derived in Appendix F. This leads to the integral:

$$\begin{aligned}\int_0^\pi \frac{\big(\cos(kh\cos\theta)-\cos(kh)\big)^2}{\sin\theta}\,d\theta &= \\ C_{\rm in}(kl)+\frac{1}{2}\cos kl\,\big[2C_{\rm in}(kl)-C_{\rm in}(2kl)\big] &+ \frac{1}{2}\sin kl\,\big[S_i(2kl)-2S_i(kl)\big]\end{aligned} \tag{16.3.6}$$

and to the radiation resistance:

$$R_{\rm peak} = \frac{\eta}{2\pi}\Big[C_{\rm in}(kl)+\frac{1}{2}\cos kl\,\big[2C_{\rm in}(kl)-C_{\rm in}(2kl)\big] + \frac{1}{2}\sin kl\,\big[S_i(2kl)-2S_i(kl)\big]\Big] \tag{16.3.7}$$

which agrees with Eq. (22.2.21) derived by a different method. The radiation resistance $R_{\rm peak}$ also determines the directivity of the dipole antenna. Using (16.3.3) for the normalized gain, we find the beam solid angle:

$$\Delta\Omega = \int_0^\pi\int_0^{2\pi} g(\theta)\,d\Omega = 2\pi c_n\int_0^\pi \frac{\big(\cos(kh\cos\theta)-\cos(kh)\big)^2}{\sin\theta}\,d\theta = 2\pi c_n\,\frac{2\pi R_{\rm peak}}{\eta}$$

which leads to the directivity-impedance relationship:

$$D_{\rm max} = \frac{4\pi}{\Delta\Omega} = \frac{1}{c_n}\,\frac{\eta}{\pi R_{\rm peak}} \tag{16.3.8}$$

The normalization constant c_n is equal to unity for a half-wave dipole; for other antenna lengths, it may be computed numerically.

The MATLAB function `dipdir` calculates c_n, the directivity $D_{\rm max}$, the angle $\theta_{\rm max}$ at which the directivity is maximum (the angle $180-\theta_{\rm max}$ also corresponds to $D_{\rm max}$), and the radiation resistance $R_{\rm peak}$. It has usage:

```
[Rpeak,Dmax,thmax,cn] = dipdir(L)        % standing-wave dipole of length L
```

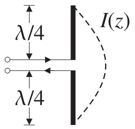
The radiation resistance is computed from Eq. (16.3.7) with the help of the sine and cosine integral functions $S_i(x)$ and $C_{\rm in}(x)$, and $D_{\rm max}$ is computed from (16.3.8).

The table below shows some representative values, with the corresponding angular patterns shown in Fig. 16.4.2.

l/λ	$R_{\rm peak}\ (\Omega)$	$D_{\rm max}$	$D_{\rm max}$ (dB)	$\theta_{\rm max}$	c_n
0.50	73.08	1.64	2.15	90.00°	1.0000
0.75	185.68	1.88	2.75	90.00°	0.3431
1.00	198.95	2.41	3.82	90.00°	0.2500
1.25	106.46	3.28	5.16	90.00°	0.3431
1.50	105.42	2.23	3.48	42.57°	0.5109
1.75	229.94	2.37	3.75	50.94°	0.2200
2.00	259.45	2.53	4.03	57.42°	0.1828
2.25	143.48	3.07	4.87	62.28°	0.2723
2.50	120.68	3.06	4.86	32.22°	0.3249

16.4 Half-Wave Dipole

The *half-wave dipole* corresponding to $l=\lambda/2$, or $kl=\pi$, is one of the most common antennas. In this case, the current distribution along the antenna takes the form:



$$\boxed{I(z)=I\cos(kz)} \qquad \text{(half-wave dipole)} \tag{16.4.1}$$

with $-\lambda/4\le z\le\lambda/4$. The normalized gain is:

$$\boxed{g(\theta)=\frac{\cos^2(0.5\pi\cos\theta)}{\sin^2\theta}} \qquad \text{(half-wave dipole gain)} \tag{16.4.2}$$

Note that the maximum does occur at $\theta=\pi/2$ and the normalization constant is $c_n=1$. Fig. 16.4.1 shows the gain in absolute and dB units. The 3-dB or half-power circle intersects the gain at an angle of $\theta_{\rm 3dB}=50.96^{\rm o}$, which leads to a half-power beam width of HPBW $=180^{\rm o}-2\theta_{\rm 3dB}=78.08^{\rm o}$, that is, somewhat narrower than the Hertzian dipole.

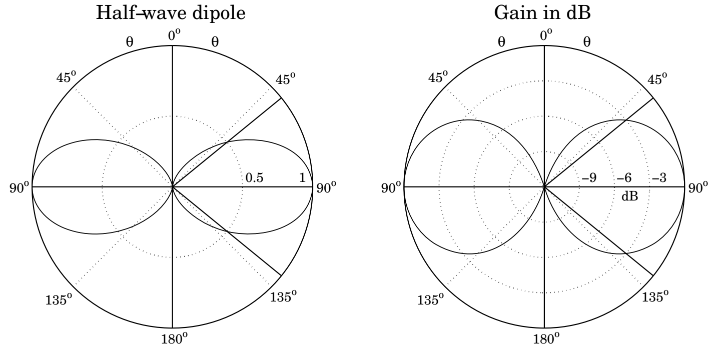


Fig. 16.4.1 Gain of half-wave dipole in absolute and dB units.

Because $\sin(kl/2)=1$, $\sin(kl)=0$, and $\cos(kl)=-1$, Eq. (16.3.7) reduces to:

$$R_{\rm in} = R_{\rm peak} = \frac{\eta}{4\pi}\,C_{\rm in}(2kl) = \frac{\eta}{4\pi}\,C_{\rm in}(2\pi) = 73.0790\ \text{ohm}$$

The directivity is found from (16.3.8) with $c_n=1$:

$$D_{\rm max} = \frac{\eta}{\pi R_{\rm peak}} = 1.64 \equiv 2.15\ \text{dB}$$

In practice, the value $R_{\rm in}=73$ ohm can be matched easily to the characteristic impedance of the feed line. For arbitrary values of the length l, the following example MATLAB code used to calculate the gain function $g(\theta)$, as well as the constant c_n and the beam solid angle, is as follows:

```
N = 200;                                    % divide [0,pi] in N angle bins
dth = pi / N;                               % bin width
th = (1:N-1) * dth;                         % excludes th=0
g = ((cos(pi*L*cos(th)) - cos(pi*L)) ./ sin(th)).^2;
th = [0, th];                               % N equally-spaced angles in [0,pi)
g = [0, g];                                 % avoids division by 0
cn = 1 / max(g);
g = cn * g;                                 % normalized to unity maximum
Om = 2 * pi * sum(g .* sin(th)) * dth;      % beam solid angle
```

where the beam solid angle is computed by the approximation to the integral:

$$\Delta\Omega = 2\pi\int_0^{\pi} g(\theta)\sin\theta\,d\theta \simeq 2\pi\sum_{i=0}^{N-1} g(\theta_i)\sin\theta_i\,\Delta\theta$$

where $\Delta\theta = \pi/N$ and $\theta_i = i\Delta\theta$, $i = 0, 1, \dots, N-1$. These operations are carried out by the functions `dipole` and `dmax`. For example, the right graph in Fig. 16.4.1 and $D_{\max}$ and $\Delta\Omega$ were generated by the MATLAB code:

```
[g, th, c] = dipole(0.5, 200);
dbp(th, g, 45, 12);
[D, Omega] = dmax(th, g);
```

Gauss-Legendre quadrature integration also produces accurate results. For example, assuming the normalization constant c_n is known, the following code fragment integrates the gain function (16.3.3) to compute the beam solid angle:

```
G = inline('(cos(pi*L*cos(th)) - cos(pi*L)).^2./sin(th).^2', 'L','th');

[w,th] = quadrs([0,pi/2,pi],32);          % use 32 points in the subintervals [0,π/2] and [π/2,π]
DOm = cn * 2*pi* w'*(G(L,th).*sin(th));   % find ΔΩ = 7.6581 for L = 0.5
```

Fig. 16.4.2 shows the gains of a variety of dipoles of different lengths. The corresponding directivities are indicated on each plot.

16.5 Monopole Antennas

A monopole antenna is half of a dipole antenna placed on top of a ground plane, as shown in Fig. 16.5.1. Assuming the plane is infinite and perfectly conducting, the monopole antenna will be equivalent to a dipole whose lower half is the *image* of the upper half.

Thus, the radiation pattern (in the upper hemisphere) will be identical to that of a dipole. Because the fields are radiated only in the upper hemisphere, the total radiated power will be half that of a dipole, and hence the corresponding radiation resistance will also be halved:

$$P_{\text{monopole}} = \frac{1}{2}P_{\text{dipole}}\,, \qquad R_{\text{monopole}} = \frac{1}{2}R_{\text{dipole}}$$

Similarly, the directivity doubles because the isotropic radiation intensity in the denominator of Eq. (15.2.2) becomes half its dipole value:

$$D_{\text{monopole}} = 2D_{\text{dipole}}$$

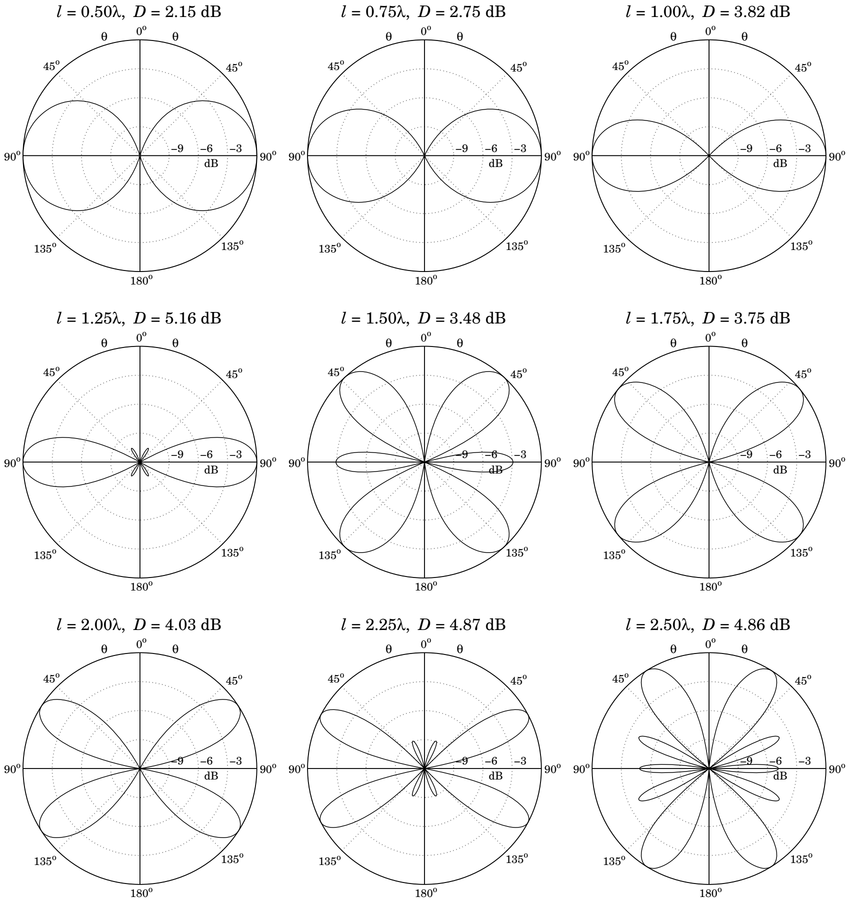


Fig. 16.4.2 Standing-wave dipole antenna patterns and directivities.

The *quarter-wave monopole* antenna whose length is $\lambda/4$ is perhaps the most widely used antenna. For AM transmitting antennas operating in the 300 m or 1 MHz band, the antenna height will be large, $\lambda/4 = 75$ m, requiring special supporting cables.

In mobile applications in the 30 cm or 1 GHz band, the antenna length will be fairly small, $\lambda/4 = 7.5$ cm. The roof of a car plays the role of the conducting plane in this case.

We note also in Fig. 16.4.2 that the $l = 1.25\lambda = 10\lambda/8$ dipole has the largest gain. It can be used as a monopole in mobile applications requiring higher gains. Such antennas are called 5/8-*wave monopoles* because their length is $l/2 = 5\lambda/8$.

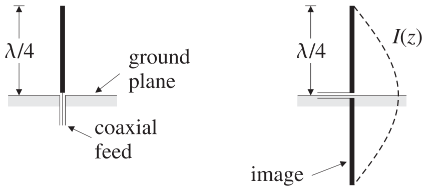


Fig. 16.5.1 Quarter-wave monopole above ground plane and the equivalent half-wave dipole.

16.6 Traveling-Wave Antennas

The standing-wave antenna current may be thought of as the linear superposition of a forward and a backward moving current. For example, the half-wave dipole current can be written in the form:

$$I(z)=I\cos(kz)=\frac{I}{2}\big(e^{-jkz}+e^{jkz}\big)$$

The backward-moving component may be eliminated by terminating the linear antenna at an appropriate matched load resistance, as shown in Fig. 16.6.1. The resulting antenna is called a *traveling-wave* antenna or a *Beverage* antenna. The current along its length has the form:

$$I(z)=Ie^{-jkz}\,,\qquad 0\le z\le l \tag{16.6.1}$$

The corresponding radiation vector becomes:

$$\mathbf{F}=\hat{\mathbf{z}}\int_0^l Ie^{-jkz'}e^{jk\cos\theta\,z'}\,dz'=\hat{\mathbf{z}}\,\frac{I}{jk}\,\frac{1-e^{-jkl(1-\cos\theta)}}{1-\cos\theta} \tag{16.6.2}$$

The transverse θ-component is:

$$F_\theta(\theta)=-F_z(\theta)\sin\theta=-\frac{I}{jk}\,\sin\theta\,\frac{1-e^{-2\pi jL(1-\cos\theta)}}{1-\cos\theta}\equiv-\frac{I}{jk}\,F(\theta) \tag{16.6.3}$$

where as before, $L=l/\lambda$ and $kl=2\pi l/\lambda=2\pi L$. The radiation intensity, given by Eq. (15.1.4) or (16.1.7), becomes now:

$$U(\theta)=\frac{\eta|I|^2}{32\pi^2}\,|F(\theta)|^2=\frac{\eta|I|^2}{8\pi^2}\left|\frac{\sin\theta\,\sin\big(\pi L(1-\cos\theta)\big)}{1-\cos\theta}\right|^2 \tag{16.6.4}$$

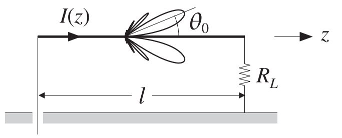


Fig. 16.6.1 Traveling-wave antenna with matched termination.

Therefore, the normalized power gain will be:

$$g(\theta)=c_n\left|\frac{\sin\theta\,\sin\big(\pi L(1-\cos\theta)\big)}{1-\cos\theta}\right|^2 \tag{16.6.5}$$

where c_n is a normalization constant. Fig. 16.6.2 shows the power gains and directivities for the cases $l=5\lambda$ and $l=10\lambda$, or $L=5$ and $L=10$.

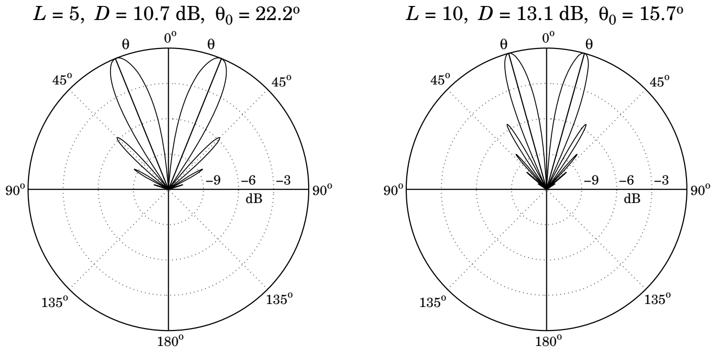


Fig. 16.6.2 Traveling-wave antenna gain examples.

The MATLAB function `travel` calculates the gain (16.6.5). For example, the left graph in Fig. 16.6.2 was generated by the MATLAB code:

```
[g, th, c, th0] = travel(5, 400);
dbp(th, g, 45, 12);
addray(90-th0,'-'); addray(90+th0,'-');
```

The longer the length l, the more the main lobes tilt towards the traveling direction of the antenna. The main lobes occur approximately at the polar angle (in radians) [5–7]:

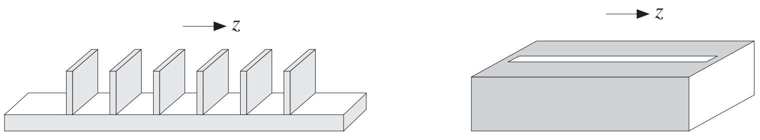
$$\boxed{\theta_0=\arccos\left(1-\frac{0.371\lambda}{l}\right)=\arccos\left(1-\frac{0.371}{L}\right)} \tag{16.6.6}$$

For the two examples of Fig. 16.6.2, this expression gives for $L=5$ and $L=10$, $\theta_0=22.2^{\rm o}$ and $\theta_0=15.7^{\rm o}$. As L increases, the angle θ_0 tends to zero.

There are other antenna structures that act as traveling-wave antennas, as shown in Fig. 16.6.3. For example, a waveguide with a long slit along its length will radiate continuously along the slit. Another example is a corrugated conducting surface along which a surface wave travels and gets radiated when it reaches the discontinuity at the end of the structure.

In all of these examples, the radiation pattern has an angular dependence similar to that of a linear antenna with a traveling-wave current of the form:

$$I(z)=Ie^{-j\beta z}=Ie^{-jpkz}\,,\qquad 0\le z\le l \tag{16.6.7}$$



currugated surface-wave antenna leaky waveguide antenna

Fig. 16.6.3 Surface-wave and leaky-wave antennas.

where β is the wavenumber along the guiding structure and $p = \beta/k = c/v_{\text{phase}}$ is the ratio of the speed of light in vacuum to the phase velocity along the guide. The corresponding radiation power pattern will now have the form:

$$g(\theta) = c_n \left| \frac{\sin\theta\, \sin\bigl(\pi L(p - \cos\theta)\bigr)}{p - \cos\theta} \right|^2 \qquad (16.6.8)$$

For long lengths L (and for $p < 1$), it peaks along the direction $\theta_0 = \arccos(p)$. Note that p can take the values: (a) $p > 1$ (slow waves), as in the case of the corrugated plane structure or the case of a Beverage antenna wrapped in a dielectric, (b) $p < 1$ (fast waves), as in the case of the leaky waveguide, where $p = \sqrt{1 - \omega_c^2/\omega^2}$, and (c) $p = 1$, for the Beverage antenna.

16.7 Vee and Rhombic Antennas

A vee antenna consists of two traveling-wave antennas forming an angle 2α with each other, as shown in Fig. 16.7.1. It may be constructed by opening up the matched ends of a transmission line at an angle of 2α (each of the terminating resistances is $R_L/2$ for a total of R_L.)

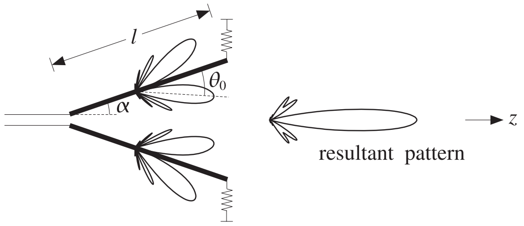


Fig. 16.7.1 Traveling-wave vee antenna with $l = 5\lambda$, $\theta_0 = 22.2^{\text{o}}$, and $\alpha = 0.85\theta_0 = 18.9^{\text{o}}$.

By choosing the angle α to be approximately equal to the main lobe angle θ_0 of Eq. (16.6.6), the two inner main lobes align with each other along the middle direction and produce a stronger main lobe, thus increasing the directivity of the antenna. The outer main lobes will also be present, but smaller.

The optimum angle α of the arms of the vee depends on the length l and is related to main lobe angle θ_0 via $\alpha = a\theta_0$, where the factor a typically falls in the range $a = 0.80$–1.00. Figure 16.7.2 shows the optimum angle factor a that corresponds to maximum directivity (in the plane of the vee) as a function of the length l.

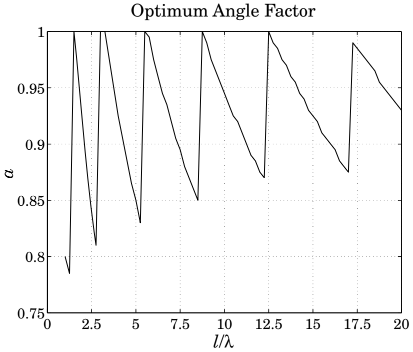


Fig. 16.7.2 Optimum angle factor as a function of antenna length.

Figure 16.7.3 shows the actual power patterns for the cases $l = 5\lambda$ and $l = 10\lambda$. The main lobe angles were $\theta_0 = 22.2^{\text{o}}$ and $\theta_0 = 15.7^{\text{o}}$. The optimum vee angles were found to be approximately (see Fig. 16.7.2), $\alpha = 0.85\theta_0 = 18.9^{\text{o}}$ and $\alpha = 0.95\theta_0 = 14.9^{\text{o}}$, in the two cases.

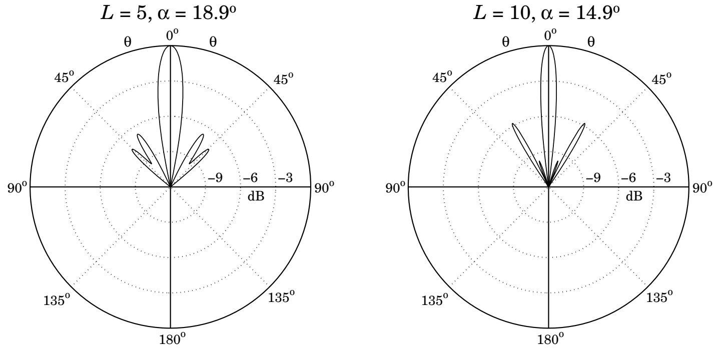


Fig. 16.7.3 Traveling-wave vee antenna gains in dB.

The combined radiation pattern can be obtained with the help of Fig. 16.7.4. Let $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$ be the two unit vectors along the two arms of the vee, and let θ_1, θ_2 be the two polar angles of the observation point P with respect to the directions $\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2$. The assumed currents along the two arms have opposite amplitudes and are:

$$I_1(z_1) = Ie^{-jkz_1}\,, \qquad I_2(z_2) = -Ie^{-jkz_2}\,, \qquad \text{for} \quad 0 \le z_1, z_2 \le l$$

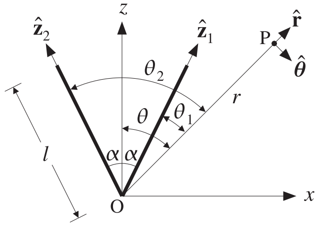


Fig. 16.7.4 Radiation vectors of traveling-wave vee antenna.

Applying the result of Eq. (16.6.2), the radiation vectors of the two arms will be:

$$\boldsymbol{F}_1 = \hat{\mathbf{z}}_1\int_0^l Ie^{-jkz_1'}e^{jk\cos\theta_1 z_1'}dz_1' = \hat{\mathbf{z}}_1\frac{I}{jk}\,\frac{1-e^{-jkl(1-\cos\theta_1)}}{1-\cos\theta_1}$$

$$\boldsymbol{F}_2 = -\hat{\mathbf{z}}_2\int_0^l Ie^{-jkz_2'}e^{jk\cos\theta_2 z_2'}dz_2' = -\hat{\mathbf{z}}_2\frac{I}{jk}\,\frac{1-e^{-jkl(1-\cos\theta_2)}}{1-\cos\theta_2}$$

Therefore, the θ-components will be as in Eq. (16.6.3):

$$F_{1\theta} = -\frac{I}{jk}F(\theta_1)\,,\qquad F_{2\theta} = \frac{I}{jk}F(\theta_2)$$

where the function $F(\theta)$ was defined in Eq. (16.6.3). From Fig. 16.7.4, we may express θ_1, θ_2 in terms of the polar angle θ with respect to the z-axis as:

$$\theta_1 = \theta - \alpha\,,\qquad \theta_2 = \theta + \alpha$$

Adding the θ-components, we obtain the resultant:

$$F_\theta = F_{1\theta} + F_{2\theta} = \frac{I}{jk}\big[F(\theta_2) - F(\theta_1)\big] = \frac{I}{jk}\big[F(\theta+\alpha) - F(\theta-\alpha)\big]$$

Thus, the radiation intensity will be:

$$U(\theta) = \frac{\eta k^2}{32\pi^2}|F_\theta(\theta)|^2 = \frac{\eta|I|^2}{32\pi^2}\big|F(\theta+\alpha) - F(\theta-\alpha)\big|^2$$

and the normalized power pattern:

$$g(\theta) = c_n\big|F(\theta+\alpha) - F(\theta-\alpha)\big|^2 \tag{16.7.1}$$

This is the gain plotted in Fig. 16.7.3 and can be computed by the MATLAB function `vee`. Finally, we consider briefly a rhombic antenna made up of two concatenated vee antennas, as shown in Fig. 16.7.5. Now the two inner main lobes of the first vee (lobes a, b) and the two outer lobes of the second vee (lobes c, d) align with each other, thus increasing the directivity of the antenna system.

The radiation vectors $\boldsymbol{F}_3$ and $\boldsymbol{F}_4$ of arms 3 and 4 may be obtained by noting that these arms are the translations of arms 1 and 2, and therefore, the radiation vectors are changed by the appropriate translational phase shift factors, as discussed in Sec. 19.2.

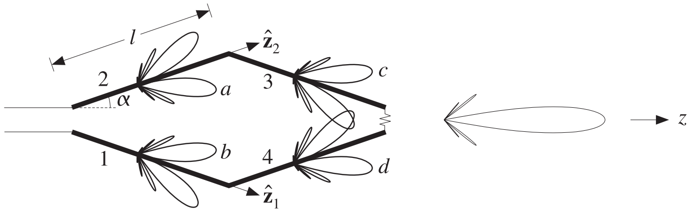


Fig. 16.7.5 Traveling-wave rhombic antenna.

Arm-3 is the translation of arm-1 by the vector $\mathbf{d}_2 = l\,\hat{\mathbf{z}}_2$ and arm-4 is the translation of arm-2 by the vector $\mathbf{d}_1 = l\,\hat{\mathbf{z}}_1$. Thus, the corresponding radiation vectors will be:

$$\boldsymbol{F}_3 = -e^{j\mathbf{k}\cdot\mathbf{d}_2}\boldsymbol{F}_1\,,\qquad \boldsymbol{F}_4 = -e^{j\mathbf{k}\cdot\mathbf{d}_1}\boldsymbol{F}_2 \tag{16.7.2}$$

where the negative signs arise because the currents in those arms have opposite signs with their parallel counterparts. The phase shift factors are:

$$e^{j\mathbf{k}\cdot\mathbf{d}_2} = e^{jkl\hat{\mathbf{r}}\cdot\hat{\mathbf{z}}_2} = e^{jkl\cos\theta_2}\,,\qquad e^{j\mathbf{k}\cdot\mathbf{d}_1} = e^{jkl\hat{\mathbf{r}}\cdot\hat{\mathbf{z}}_1} = e^{jkl\cos\theta_1}$$

It follows that the θ-components of $\boldsymbol{F}_3$ and $\boldsymbol{F}_4$ are:

$$F_{3\theta} = -e^{jkl\cos\theta_2}F_{1\theta} = \frac{I}{jk}e^{jkl\cos\theta_2}F(\theta_1)$$

$$F_{4\theta} = -e^{jkl\cos\theta_1}F_{2\theta} = -\frac{I}{jk}e^{jkl\cos\theta_1}F(\theta_2)$$

Thus, the resultant θ-component will be:

$$F_\theta = F_{1\theta} + F_{2\theta} + F_{3\theta} + F_{4\theta} = \frac{I}{jk}\big[F(\theta_2) - F(\theta_1) + e^{jkl\cos\theta_2}F(\theta_1) - e^{jkl\cos\theta_1}F(\theta_2)\big]$$

The corresponding normalized power pattern will be:

$$g(\theta) = c_n\big|F(\theta+\alpha) - F(\theta-\alpha) + e^{jkl\cos(\theta+\alpha)}F(\theta-\alpha) - e^{jkl\cos(\theta-\alpha)}F(\theta+\alpha)\big|^2$$

Figure 16.7.6 shows the power gain $g(\theta)$ for the cases $L = 5$ and $L = 10$. The optimum vee angle in both cases was found to be $\alpha = \theta_0$, that is, $\alpha = 22.2^{\circ}$ and $\alpha = 15.7^{\circ}$. The function `rhombic` may be used to evaluate this expression.

16.8 Loop Antennas

Figure 16.8.1 shows a circular and a square loop antenna. The feed points are not shown. The main oversimplifying assumption here is that the current is constant around the loop. We will mainly consider the case when the dimension of the loop (e.g., its circumference) is small relative to the wavelength.

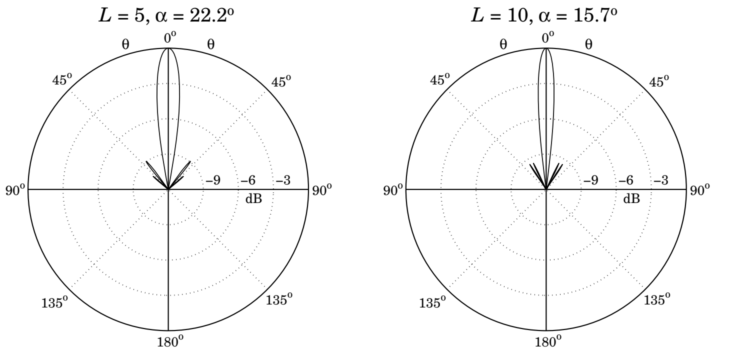


Fig. 16.7.6 Rhombic antenna gains in dB.

For such small loops, the radiation pattern turns out to be *independent* of the shape of the loop and the radiation vector takes the simple form:

$$\boldsymbol{F} = j\,\mathbf{m}\times\boldsymbol{k} \tag{16.8.1}$$

where $\mathbf{m}$ is the loop's magnetic moment defined with respect to Fig. 16.8.1 as follows:

$$\mathbf{m} = \hat{\mathbf{z}}\,IS\,, \qquad \text{(magnetic moment)} \tag{16.8.2}$$

where S is the area of the loop. Writing $\boldsymbol{k} = k\,\hat{\mathbf{r}}$ and noting that $\hat{\mathbf{z}}\times\hat{\mathbf{r}} = \hat{\boldsymbol{\phi}}\sin\theta$, we have:

$$\boldsymbol{F} = j\,\mathbf{m}\times\boldsymbol{k} = jmk\sin\theta\,\hat{\boldsymbol{\phi}} \equiv F_\phi(\theta)\,\hat{\boldsymbol{\phi}} \tag{16.8.3}$$

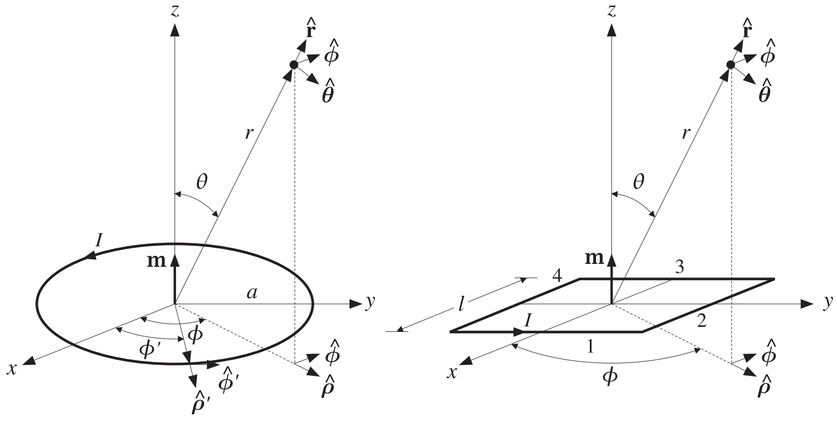


Fig. 16.8.1 Circular and square loop antennas.

Thus, $\boldsymbol{F}$ is fully transverse to $\hat{\mathbf{r}}$, so that $\boldsymbol{F}_\perp = \boldsymbol{F}$. It follows from Eq. (14.10.4) that the produced radiation fields will be:

$$\begin{aligned}\boldsymbol{E} &= \hat{\boldsymbol{\phi}}E_\phi = -jk\eta\,\frac{e^{-jkr}}{4\pi r}F_\phi\,\hat{\boldsymbol{\phi}} = \eta mk^2\sin\theta\,\frac{e^{-jkr}}{4\pi r}\,\hat{\boldsymbol{\phi}}\\ \boldsymbol{H} &= \hat{\boldsymbol{\theta}}H_\theta = jk\,\frac{e^{-jkr}}{4\pi r}F_\phi\,\hat{\boldsymbol{\theta}} = -mk^2\sin\theta\,\frac{e^{-jkr}}{4\pi r}\,\hat{\boldsymbol{\theta}}\end{aligned} \tag{16.8.4}$$

The radiation intensity of Eq. (15.1.4) is in this case:

$$\boxed{U(\theta,\phi) = \frac{\eta k^2}{32\pi^2}\,|F_\phi|^2 = \frac{\eta k^4|m|^2}{32\pi^2}\,\sin^2\theta} \qquad \text{(loop intensity)} \tag{16.8.5}$$

Thus, it has the same $\sin^2\theta$ angular dependence, normalized power gain, and directivity as the Hertzian dipole. We may call such small loop antennas "Hertzian loops", referring to their infinitesimal size. The total radiated power can be computed as in Sec. 16.2. We have:

$$P_{\text{rad}} = U_{\max}\Delta\Omega = \frac{\eta k^4|m|^2}{32\pi^2}\,\frac{8\pi}{3} = \frac{\eta k^4|m|^2}{12\pi}$$

Replacing m by IS, we may obtain the loop's radiation resistance from the definition:

$$P_{\text{rad}} = \frac{1}{2}R_{\text{rad}}|I|^2 = \frac{\eta k^4|IS|^2}{12\pi} \qquad \Rightarrow \qquad R_{\text{rad}} = \frac{\eta k^4 S^2}{6\pi}$$

Comparing Eq. (16.8.4) to the Hertzian dipole, the loop's electric field is in the ϕ-direction, whereas the Hertzian dipole's is in the θ-direction. The relative amplitudes of the electric fields are:

$$\frac{E_\theta^{\text{dipole}}}{E_\phi^{\text{loop}}} = j\,\frac{Il}{mk}$$

If we choose $Il = mk$, then the electric fields are off by a 90°-degree phase. If such a Hertzian dipole and loop are placed at the origin, the produced net electric field will be *circularly* polarized. We note finally that the loop may have several turns, thus increasing its radiation resistance and radiated power. For a loop with n turns, we must make the replacement $m \to nm$.

16.9 Circular Loops

Next, we consider the circular loop in more detail, and derive Eq. (16.8.3). Assuming an infinitely thin wire loop of radius a, the assumed current density can be expressed in cylindrical coordinates as in Eq. (16.1.3):

$$\boldsymbol{J}(\mathbf{r}') = I\,\hat{\boldsymbol{\phi}}'\,\delta(\rho'-a)\,\delta(z')$$

The radiation vector will be:

$$\boldsymbol{F} = \int_V \boldsymbol{J}(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,d^3\mathbf{r}' = \int I\,\hat{\boldsymbol{\phi}}'\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,\delta(\rho'-a)\,\delta(z')\,\rho'\,d\rho'\,d\phi'\,dz' \tag{16.9.1}$$

Using Eq. (14.8.2), we have:

$$\begin{aligned}\boldsymbol{k}\cdot\mathbf{r}' &= k(\hat{\mathbf{z}}\cos\theta+\hat{\boldsymbol{\rho}}\sin\theta)\cdot(z'\hat{\mathbf{z}}'+\rho'\hat{\boldsymbol{\rho}}')\\ &= kz'\cos\theta+k\rho'\sin\theta\,(\hat{\boldsymbol{\rho}}'\cdot\hat{\boldsymbol{\rho}})\\ &= kz'\cos\theta+k\rho'\sin\theta\cos(\phi'-\phi)\end{aligned}$$

where we set $\hat{\boldsymbol{\rho}}'\cdot\hat{\boldsymbol{\rho}}=\cos(\phi'-\phi)$, as seen in Fig. 16.8.1. The integration in Eq. (16.9.1) confines $\mathbf{r}'$ to the xy-plane and sets $\rho'=a$ and $z'=0$. Thus, we have in the integrand:

$$\boldsymbol{k}\cdot\mathbf{r}'=ka\sin\theta\cos(\phi'-\phi)$$

Then, the radiation vector (16.9.1) becomes:

$$\boldsymbol{F}=Ia\int_0^{2\pi}\hat{\boldsymbol{\phi}}'e^{jka\sin\theta\cos(\phi'-\phi)}d\phi' \tag{16.9.2}$$

We note in Fig. 16.8.1 that the unit vector $\hat{\boldsymbol{\phi}}'$ varies in direction with ϕ'. Therefore, it proves convenient to express it in terms of the unit vectors $\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\rho}}$ of the fixed observation point P. Resolving $\hat{\boldsymbol{\phi}}'$ into the directions $\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\rho}}$, we have:

$$\hat{\boldsymbol{\phi}}'=\hat{\boldsymbol{\phi}}\cos(\phi'-\phi)-\hat{\boldsymbol{\rho}}\sin(\phi'-\phi)$$

Changing integration variables from ϕ' to $\psi=\phi'-\phi$, we write Eq. (16.9.2) as:

$$\boldsymbol{F}=Ia\int_0^{2\pi}(\hat{\boldsymbol{\phi}}\cos\psi-\hat{\boldsymbol{\rho}}\sin\psi)e^{jka\sin\theta\cos\psi}d\psi$$

The second term is odd in ψ and vanishes. Thus,

$$\boldsymbol{F}=Ia\,\hat{\boldsymbol{\phi}}\int_0^{2\pi}\cos\psi e^{jka\sin\theta\cos\psi}d\psi \tag{16.9.3}$$

Using the integral representation of the Bessel function $J_1(x)$,

$$J_1(x)=\frac{1}{2\pi j}\int_0^{2\pi}\cos\psi\,e^{jx\cos\psi}d\psi$$

we may replace the ψ-integral by $2\pi jJ_1(ka\sin\theta)$ and write Eq. (16.9.3) as:

$$\boldsymbol{F}=F_\phi\hat{\boldsymbol{\phi}}=2\pi j\,Ia\,J_1(ka\sin\theta)\hat{\boldsymbol{\phi}} \tag{16.9.4}$$

This gives the radiation vector for any loop radius. If the loop is electrically small, that is, $ka\ll 1$, we may use the first-order approximation $J_1(x)\simeq x/2$, to get

$$\boldsymbol{F}=F_\phi\hat{\boldsymbol{\phi}}=2\pi j\,Ia\,\frac{1}{2}ka\sin\theta\,\hat{\boldsymbol{\phi}}=jI\pi a^2k\sin\theta\,\hat{\boldsymbol{\phi}} \tag{16.9.5}$$

which agrees with Eq. (16.8.3), with $m=IS=I\pi a^2$.

16.10 Square Loops

The square loop of Fig. 16.8.1 may be thought of as four separate linear antennas representing the four sides. Assuming that each side is a Hertzian dipole and that the sides are at distances $\pm l/2$ from the origin, we can write the current densities of the sides 1, 2, 3, 4 as follows:

$$\begin{aligned}\boldsymbol{J}_1(\mathbf{r}) &= \hat{\mathbf{y}}Il\,\delta(x-l/2)\delta(y)\delta(z)\\ \boldsymbol{J}_2(\mathbf{r}) &= -\hat{\mathbf{x}}Il\,\delta(x)\delta(y-l/2)\delta(z)\\ \boldsymbol{J}_3(\mathbf{r}) &= -\hat{\mathbf{y}}Il\,\delta(x+l/2)\delta(y)\delta(z)\\ \boldsymbol{J}_4(\mathbf{r}) &= \hat{\mathbf{x}}Il\,\delta(x)\delta(y+l/2)\delta(z)\end{aligned}$$

The currents on the parallel sides 1 and 3 combine to give:

$$\boldsymbol{J}_1(\mathbf{r})+\boldsymbol{J}_3(\mathbf{r})=-Il^2\,\hat{\mathbf{y}}\left[\frac{\delta(x+l/2)-\delta(x-l/2)}{l}\right]\delta(y)\delta(z)$$

where we multiplied and divided by a factor of l. In the limit of small l, we may replace the quantity in the bracket by the derivative $\delta'(x)$ of the delta function $\delta(x)$:

$$\boldsymbol{J}_1(\mathbf{r})+\boldsymbol{J}_3(\mathbf{r})=-Il^2\,\hat{\mathbf{y}}\,\delta'(x)\delta(y)\delta(z)$$

Similarly, we find for sides 2 and 4:

$$\boldsymbol{J}_2(\mathbf{r})+\boldsymbol{J}_4(\mathbf{r})=Il^2\,\hat{\mathbf{x}}\,\delta(x)\delta'(y)\delta(z)$$

Thus, the net current density of all sides is:

$$\boldsymbol{J}(\mathbf{r})=Il^2\big[\hat{\mathbf{x}}\,\delta(x)\delta'(y)-\hat{\mathbf{y}}\,\delta'(x)\delta(y)\big]\delta(z) \tag{16.10.1}$$

The corresponding radiation vector will be:

$$\boldsymbol{F}=Il^2\int\big[\hat{\mathbf{x}}\,\delta(x')\delta'(y')-\hat{\mathbf{y}}\,\delta'(x')\delta(y')\big]\delta(z')e^{j(k_xx'+k_yy'+k_zz')}dx'dy'dz'$$

The delta-function integrations can be done easily yielding:

$$\boldsymbol{F}=Il^2(-jk_y\hat{\mathbf{x}}+jk_x\hat{\mathbf{y}})$$

Using Eq. (16.1.4), we obtain

$$\boldsymbol{F}=jIl^2k\sin\theta(-\hat{\mathbf{x}}\sin\phi+\hat{\mathbf{y}}\cos\phi)=jIl^2k\sin\theta\,\hat{\boldsymbol{\phi}} \tag{16.10.2}$$

which agrees with Eq. (16.8.3), with $m=IS=Il^2$.

16.11 Dipole and Quadrupole Radiation

The radiation vector $\boldsymbol{F}$ of a current/charge distribution can be evaluated approximately by expanding the exponential $e^{j\boldsymbol{k}\cdot\mathbf{r}'}$ to successive powers of $\boldsymbol{k}$:

$$\begin{aligned}\boldsymbol{F} &= \int_V \boldsymbol{J}(\mathbf{r}')\,e^{j\boldsymbol{k}\cdot\mathbf{r}'}d^3\mathbf{r}' = \int_V \big[1 + j\boldsymbol{k}\cdot\mathbf{r}' + \frac{1}{2!}(j\boldsymbol{k}\cdot\mathbf{r}')^2 + \cdots\big]\boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' \\ &= \underbrace{\int_V \boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}'}_{\text{elec. dipole}} + \underbrace{\int_V j(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}'}_{\text{magn. dipole \& elec. quadrupole}} + \cdots\end{aligned} \tag{16.11.1}$$

The first term is the *electric dipole* radiation term and corresponds to the Hertzian dipole antenna. The second term incorporates both the *magnetic dipole* (corresponding to a Hertzian loop antenna) and the *electric quadrupole* terms.

Higher multipoles arise from the higher-order terms in the above expansion. A systematic discussion of all multipole radiation terms requires the use of spherical harmonics.

Keeping only a few terms in the above expansion is a good approximation to $\boldsymbol{F}$ provided $kr' \ll 1$, or $l \ll \lambda$, where l is the typical dimension of the current source. In general, any radiating system will emit radiation of various multipole types.

The *electric dipole* and *electric quadrupole moments* of a charge distribution are defined in terms of the following first- and second-order moments of the charge density:

$$\mathbf{p} = \int_V \mathbf{r}'\rho(\mathbf{r}')\,d^3\mathbf{r}' \qquad \text{(electric dipole moment)} \tag{16.11.2}$$

$$D_{ij} = \int_V r_i' r_j' \rho(\mathbf{r}')\,d^3\mathbf{r}' \qquad \text{(electric quadrupole moment)} \tag{16.11.3}$$

The identity of Problem 14.2 is useful here in manipulating the successive expansion terms of $\boldsymbol{F}$. Applying the identity with the two choices: $g(\mathbf{r}') = r_i'$ and $g(\mathbf{r}') = r_i' r_j'$, we obtain the relationships:

$$\begin{aligned}\int_V J_i\,d^3\mathbf{r}' &= j\omega\int_V r_i'\rho(\mathbf{r}')\,d^3\mathbf{r}' = j\omega p_i \\ \int_V (r_i'J_j + r_j'J_i)\,d^3\mathbf{r}' &= j\omega\int_V r_i'r_j'\rho(\mathbf{r}')\,d^3\mathbf{r}' = j\omega D_{ij}\end{aligned} \tag{16.11.4}$$

Thus, the lowest-order term in Eq. (16.11.1) is the electric dipole:

$$\int_V \boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' = j\omega\,\mathbf{p} = \boldsymbol{F}_{\text{el}}$$

In the second term of Eq. (16.11.1), we may apply the vectorial identity:

$$(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J} = \frac{1}{2}(\mathbf{r}'\times\boldsymbol{J})\times\boldsymbol{k} + \frac{1}{2}\big[(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J} + (\boldsymbol{k}\cdot\boldsymbol{J})\mathbf{r}'\big]$$

and in integrated form:

$$\int_V(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J}\,d^3\mathbf{r}' = \frac{1}{2}\int_V(\mathbf{r}'\times\boldsymbol{J})\times\boldsymbol{k}\,d^3\mathbf{r}' + \frac{1}{2}\int_V\big[(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J} + (\boldsymbol{k}\cdot\boldsymbol{J})\mathbf{r}'\big]\,d^3\mathbf{r}' \tag{16.11.5}$$

The *magnetic moment* of a current distribution is defined in general by

$$\mathbf{m} = \frac{1}{2}\int_V \mathbf{r}'\times\boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' \qquad \text{(magnetic moment)} \tag{16.11.6}$$

Therefore, the first term in Eq. (16.11.5) may be written as $\mathbf{m}\times\boldsymbol{k}$. With the help of the second identity of Eq. (16.11.4), the last term of (16.11.5) may be written in terms of the quadrupole matrix D acting on the vector $\boldsymbol{k}$. We have then for the second term in the expansion (16.11.1):

$$\int_V j(\boldsymbol{k}\cdot\mathbf{r}')\boldsymbol{J}\,d^3\mathbf{r}' = j\,\mathbf{m}\times\boldsymbol{k} - \frac{1}{2}\omega D\boldsymbol{k} = \boldsymbol{F}_{\text{mag}} + \boldsymbol{F}_{\text{quad}} \tag{16.11.7}$$

Thus, the three lowest-order terms of $\boldsymbol{F}$ are:

$$\boxed{\boldsymbol{F} = \boldsymbol{F}_{\text{el}} + \boldsymbol{F}_{\text{mag}} + \boldsymbol{F}_{\text{quad}} = j\omega\,\mathbf{p} + j\,\mathbf{m}\times\boldsymbol{k} - \frac{1}{2}\omega D\boldsymbol{k}} \tag{16.11.8}$$

We briefly discuss each term. For a Hertzian dipole antenna with $\boldsymbol{J}(\mathbf{r}') = \hat{\mathbf{z}}\,Il\,\delta^3(\mathbf{r}')$, only the first term of (16.11.8) is non-zero and is the same as that of Sec. 16.2:

$$\boldsymbol{F}_{\text{el}} = \int_V \boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' = \hat{\mathbf{z}}\,Il = j\omega\,\mathbf{p}$$

The relationship $Il = j\omega p$ may be understood by thinking of the Hertzian dipole as two opposite time-varying charges $\pm q$ separated by a distance l (along the z-direction), so that $p = ql$. It follows that $j\omega p = \dot{p} = \dot{q}l = Il$.

The result $p = ql$ may also be applied to the case of an accelerated charge. Now q is constant but l varies with time. We have $\dot{p} = q\dot{l} = qv$ and $\ddot{p} = q\dot{v} = qa$, where a is the acceleration $a = \dot{v}$. For harmonic time dependence, we have $(j\omega)^2 p = qa$. The total radiated power from a dipole was obtained in Eq. (16.2.2). Setting $k^2|Il|^2 = k^2|qv|^2 = q^2\omega^2|v|^2/c^2 = q^2|a|^2/c^2$, we can rewrite Eq. (16.2.2) in the form:

$$P = \frac{\eta q^2|a|^2}{12\pi c^2} = \frac{\eta q^2 a_{\text{rms}}^2}{6\pi c^2}$$

where $a_{\text{rms}} = |a|/\sqrt{2}$ is the rms value of the acceleration. This is Larmor's classical expression for the radiated power from a nonrelativistic accelerated charge.

For a Hertzian loop, only the magnetic moment term is present in $\boldsymbol{F}$. We may verify the result that $\mathbf{m} = \hat{\mathbf{z}}\,IS$ using the definition (16.11.6). Indeed, for a circular loop:

$$\mathbf{m} = \frac{1}{2}\int \mathbf{r}'\times\big[I\hat{\boldsymbol{\phi}}'\delta(\rho' - a)\delta(z')\big]\rho'\,d\rho'\,d\phi'\,dz'$$

The integrations over z' and ρ' force $z' = 0$ and $\rho' = a$, and therefore, $\mathbf{r}' = a\hat{\boldsymbol{\rho}}'$. Noting that $\hat{\boldsymbol{\rho}}'\times\hat{\boldsymbol{\phi}}' = \hat{\mathbf{z}}$ and that the ϕ'-integration contributes a factor of 2π, we obtain:

$$\mathbf{m} = \frac{1}{2}a\,\hat{\boldsymbol{\rho}}'\times\hat{\boldsymbol{\phi}}'\,Ia\,2\pi = \hat{\mathbf{z}}\,I(\pi a^2)$$

Similarly, inserting Eq. (16.10.1) into (16.11.6), we find for the square loop:

$$\mathbf{m} = \frac{1}{2}\int(x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}})\times\big[Il^2\big(\hat{\mathbf{x}}\,\delta(x)\,\delta'(y) - \hat{\mathbf{y}}\,\delta'(x)\,\delta(y)\big)\delta(z)\big]dx\,dy\,dz = \hat{\mathbf{z}}\,Il^2$$

For the electric quadrupole term, the matrix D is sometimes replaced by its *traceless* version defined by

$$Q_{ij} = 3D_{ij} - \delta_{ij}\mathrm{tr}(D) = \int_V \big(3r'_i r'_j - \delta_{ij}\,\mathbf{r}'\cdot\mathbf{r}'\big)\rho(\mathbf{r}')\,d^3\mathbf{r}' \quad\Rightarrow\quad Q = 3D - I\,\mathrm{tr}(D)$$

so that $\mathrm{tr}(Q) = 0$. In this case, the vector $D\boldsymbol{k}$ may be expressed as

$$D\boldsymbol{k} = \frac{1}{3}\,Q\boldsymbol{k} + \frac{1}{3}\,\mathrm{tr}(D)\;\boldsymbol{k}$$

The second term may be ignored because it does not contribute to the radiation fields, which depend only on the part of $\boldsymbol{F}$ transverse to $\boldsymbol{k}$. Thus, without loss of generality we may also write:

$$\boldsymbol{F} = j\omega\,\mathbf{p} + j\,\mathbf{m}\times\boldsymbol{k} - \frac{1}{6}\,\omega Q\boldsymbol{k}$$

The electric and magnetic dipoles have angular gain patterns that are identical to the Hertzian dipole and Hertzian loop antennas, that is, $\sin^2\theta$. The quadrupole term, on the other hand, can have a complicated angular pattern as can be seen by expressing the vector $Q\boldsymbol{k} = kQ\hat{\mathbf{r}}$ explicitly in terms of the angles θ,ϕ:

$$Q\hat{\mathbf{r}} = \begin{bmatrix} Q_{xx} & Q_{xy} & Q_{xz} \\ Q_{yx} & Q_{yy} & Q_{yz} \\ Q_{zx} & Q_{zy} & Q_{zz} \end{bmatrix}\begin{bmatrix} \sin\theta\,\cos\phi \\ \sin\theta\,\sin\phi \\ \cos\theta \end{bmatrix}$$

16.12 Problems

16.1 *Computer Experiment—Dipoles.* Reproduce the results and graphs of Fig. 16.4.2, and calculate the corresponding directivities in dB.

16.2 Derive Eq. (16.3.7) for the input resistance of a dipole antenna.

16.3 Derive Eq. (16.6.6) for the tilt angle of a traveling wave antenna by reducing the problem to that of finding the maximum of the function $\sin^2(\pi x)/x$ in the interval $[0,1]$.

16.4 *Computer Experiment–Traveling Wave Antennas.* Reproduce the results and graphs of Fig. 16.6.2.

17

Radiation from Apertures

17.1 Field Equivalence Principle

The radiation fields from aperture antennas, such as slots, open-ended waveguides, horns, reflector and lens antennas, are determined from the knowledge of the fields *over* the aperture of the antenna.

The aperture fields become the *sources* of the radiated fields at large distances. This is a variation of the Huygens-Fresnel principle, which states that the points on each wavefront become the sources of secondary spherical waves propagating outwards and whose superposition generates the next wavefront.

Let $\boldsymbol{E}_a, \boldsymbol{H}_a$ be the *tangential fields* over an aperture A, as shown in Fig. 17.1.1. These fields are assumed to be known and are produced by the sources to the left of the screen. The problem is to determine the radiated fields $\boldsymbol{E}(\mathbf{r}), \boldsymbol{H}(\mathbf{r})$ at some far observation point.

The radiated fields can be computed with the help of the *field equivalence principle* [1112–1118,1167], which states that the aperture fields may be replaced by *equivalent electric and magnetic surface currents*, whose radiated fields can then be calculated using the techniques of Sec. 14.10. The equivalent surface currents are:

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= \hat{\mathbf{n}} \times \boldsymbol{H}_a \\ \boldsymbol{J}_{ms} &= -\hat{\mathbf{n}} \times \boldsymbol{E}_a \end{aligned}} \quad \begin{aligned} &\text{(electric surface current)} \\ &\text{(magnetic surface current)} \end{aligned} \tag{17.1.1}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the surface and on the side of the radiated fields.

Thus, it becomes necessary to consider Maxwell's equations in the presence of magnetic currents and derive the radiation fields from such currents.

The screen in Fig. 17.1.1 is an arbitrary infinite surface over which the tangential fields are assumed to be zero. This assumption is not necessarily consistent with the radiated field solutions, that is, Eqs. (17.4.9). A consistent calculation of the fields to the right of the aperture plane requires knowledge of the fields over the entire aperture plane (screen plus aperture.)

However, for large apertures (with typical dimension much greater than a wavelength), the approximation of using the fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ only over the aperture to calculate the radiation patterns is fairly adequate, especially in predicting the main-lobe behavior of the patterns.

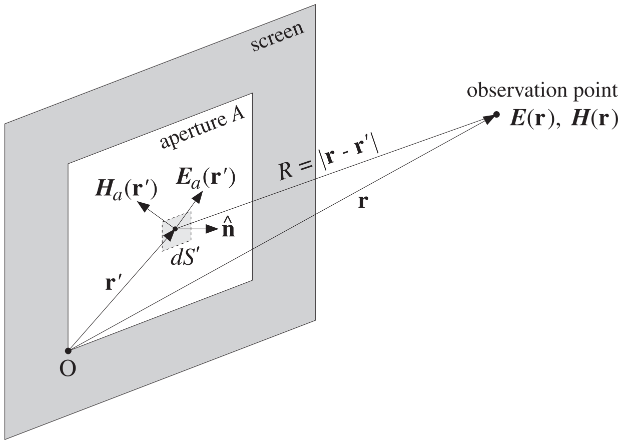


Fig. 17.1.1 Radiated fields from an aperture.

The screen can also be a perfectly conducting surface, such as a ground plane, on which the aperture opening has been cut. In reflector antennas, the aperture itself is not an opening, but rather a reflecting surface. Fig. 17.1.2 depicts some examples of screens and apertures: (a) an open-ended waveguide over an infinite ground plane, (b) an open-ended waveguide radiating into free space, and (c) a reflector antenna.

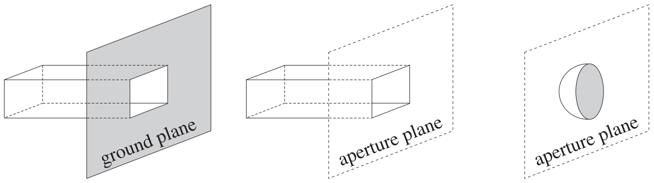


Fig. 17.1.2 Examples of aperture planes.

There are two alternative forms of the field equivalence principle, which may be used when only one of the aperture fields $\boldsymbol{E}_a$ or $\boldsymbol{H}_a$ is available. They are:

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= 0 \\ \boldsymbol{J}_{ms} &= -2\,(\hat{\mathbf{n}}\times\boldsymbol{E}_a) \end{aligned}} \qquad \text{(perfect magnetic conductor)} \tag{17.1.2}$$

$$\boxed{\begin{aligned} \boldsymbol{J}_s &= 2\,(\hat{\mathbf{n}}\times\boldsymbol{H}_a) \\ \boldsymbol{J}_{ms} &= 0 \end{aligned}} \qquad \text{(perfect electric conductor)} \tag{17.1.3}$$

They are appropriate when the screen is a *perfect electric conductor* (PEC) on which $\boldsymbol{E}_a = 0$, or when it is a *perfect magnetic conductor* (PMC) on which $\boldsymbol{H}_a = 0$.

Using image theory, the perfect electric (magnetic) conducting screen can be eliminated and replaced by an image magnetic (electric) surface current, doubling its value over the aperture. The image field causes the total tangential electric (magnetic) field to vanish over the screen.

If the tangential fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ were known over the entire aperture plane (screen plus aperture), the three versions of the equivalence principle would generate the same radiated fields. But because we consider $\boldsymbol{E}_a, \boldsymbol{H}_a$ only over the aperture, the three versions give slightly different results.

In the case of a perfectly conducting screen, the calculated radiation fields (17.4.10) using the equivalent currents (17.1.2) are consistent with the boundary conditions on the screen.

17.2 Magnetic Currents and Duality

Next, we consider the solution of Maxwell's equations driven by the ordinary electric charge and current densities $\rho, \boldsymbol{J}$, and in addition, by the magnetic charge and current densities $\rho_m, \boldsymbol{J}_m$.

Although $\rho_m, \boldsymbol{J}_m$ are fictitious, the solution of this problem will allow us to identify the equivalent magnetic currents to be used in aperture problems, and thus, establish the field equivalence principle. The generalized form of Maxwell's equations is:

$$\boxed{\begin{aligned} \nabla\times\boldsymbol{H} &= \boldsymbol{J} + j\omega\epsilon\boldsymbol{E} \\ \nabla\cdot\boldsymbol{E} &= \frac{1}{\epsilon}\,\rho \\ \nabla\times\boldsymbol{E} &= -\boldsymbol{J}_m - j\omega\mu\boldsymbol{H} \\ \nabla\cdot\boldsymbol{H} &= \frac{1}{\mu}\,\rho_m \end{aligned}} \tag{17.2.1}$$

There is now complete symmetry, or duality, between the electric and the magnetic quantities. In fact, it can be verified easily that the following *duality transformation* leaves the set of four equations *invariant*:

$$\boxed{\begin{array}{lll} \boldsymbol{E} \longrightarrow \boldsymbol{H} & \boldsymbol{J} \longrightarrow \boldsymbol{J}_m & \boldsymbol{A} \longrightarrow \boldsymbol{A}_m \\ \boldsymbol{H} \longrightarrow -\boldsymbol{E} & \rho \longrightarrow \rho_m & \varphi \longrightarrow \varphi_m \\ \epsilon \longrightarrow \mu & \boldsymbol{J}_m \longrightarrow -\boldsymbol{J} & \boldsymbol{A}_m \longrightarrow -\boldsymbol{A} \\ \mu \longrightarrow \epsilon & \rho_m \longrightarrow -\rho & \varphi_m \longrightarrow -\varphi \end{array}} \qquad \text{(duality)} \tag{17.2.2}$$

where $\varphi, \boldsymbol{A}$ and $\varphi_m, \boldsymbol{A}_m$ are the corresponding scalar and vector potentials introduced below. These transformations can be recognized as a special case (for $\alpha = \pi/2$) of the following *duality rotations*, which also leave Maxwell's equations invariant:

$$\begin{bmatrix} \boldsymbol{E}' & \eta\boldsymbol{J}' & \eta\rho' \\ \eta\boldsymbol{H}' & \boldsymbol{J}'_m & \rho'_m \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \boldsymbol{E} & \eta\boldsymbol{J} & \eta\rho \\ \eta\boldsymbol{H} & \boldsymbol{J}_m & \rho_m \end{bmatrix} \tag{17.2.3}$$

Under the duality transformations (17.2.2), the first two of Eqs. (17.2.1) transform into the last two, and conversely, the last two transform into the first two.

A useful consequence of duality is that if one has obtained expressions for the electric field $\boldsymbol{E}$, then by applying a duality transformation one can generate expressions for the magnetic field $\boldsymbol{H}$. We will see examples of this property shortly.

The solution of Eq. (17.2.1) is obtained in terms of the usual scalar and vector potentials $\varphi, \boldsymbol{A}$, as well as two new potentials $\varphi_m, \boldsymbol{A}_m$ of the magnetic type:

$$\begin{aligned} \boldsymbol{E} &= -\nabla\varphi - j\omega\boldsymbol{A} - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= -\nabla\varphi_m - j\omega\boldsymbol{A}_m + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned} \tag{17.2.4}$$

The expression for $\boldsymbol{H}$ can be derived from that of $\boldsymbol{E}$ by a duality transformation of the form (17.2.2). The scalar and vector potentials satisfy the Lorenz conditions and Helmholtz wave equations:

$$\begin{aligned} &\nabla\cdot\boldsymbol{A} + j\omega\epsilon\mu\,\varphi = 0 & & \nabla\cdot\boldsymbol{A}_m + j\omega\epsilon\mu\,\varphi_m = 0 \\ &\nabla^2\varphi + k^2\varphi = -\frac{\rho}{\epsilon} & \text{and}\quad & \nabla^2\varphi_m + k^2\varphi_m = -\frac{\rho_m}{\mu} \\ &\nabla^2\boldsymbol{A} + k^2\boldsymbol{A} = -\mu\boldsymbol{J} & & \nabla^2\boldsymbol{A}_m + k^2\boldsymbol{A}_m = -\epsilon\boldsymbol{J}_m \end{aligned} \tag{17.2.5}$$

The solutions of the Helmholtz equations are given in terms of $G(\mathbf{r}-\mathbf{r}') = \dfrac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}$:

$$\begin{aligned} \varphi(\mathbf{r}) &= \int_V \frac{1}{\epsilon}\rho(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV', & \varphi_m(\mathbf{r}) &= \int_V \frac{1}{\mu}\rho_m(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV' \\ \boldsymbol{A}(\mathbf{r}) &= \int_V \mu\boldsymbol{J}(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV', & \boldsymbol{A}_m(\mathbf{r}) &= \int_V \epsilon\boldsymbol{J}_m(\mathbf{r}')G(\mathbf{r}-\mathbf{r}')\,dV' \end{aligned} \tag{17.2.6}$$

where V is the volume over which the charge and current densities are nonzero. The observation point $\mathbf{r}$ is taken to be outside this volume. Using the Lorenz conditions, the scalar potentials may be eliminated in favor of the vector potentials, resulting in the alternative expressions for Eq. (17.2.4):

$$\boxed{\begin{aligned} \boldsymbol{E} &= \frac{1}{j\omega\mu\epsilon}\big[\nabla(\nabla\cdot\boldsymbol{A}) + k^2\boldsymbol{A}\big] - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= \frac{1}{j\omega\mu\epsilon}\big[\nabla(\nabla\cdot\boldsymbol{A}_m) + k^2\boldsymbol{A}_m\big] + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned}} \tag{17.2.7}$$

These may also be written in the form of Eq. (14.3.9):

$$\boxed{\begin{aligned} \boldsymbol{E} &= \frac{1}{j\omega\mu\epsilon}\big[\nabla\times(\nabla\times\boldsymbol{A}) - \mu\boldsymbol{J}\big] - \frac{1}{\epsilon}\nabla\times\boldsymbol{A}_m \\ \boldsymbol{H} &= \frac{1}{j\omega\mu\epsilon}\big[\nabla\times(\nabla\times\boldsymbol{A}_m) - \epsilon\boldsymbol{J}_m\big] + \frac{1}{\mu}\nabla\times\boldsymbol{A} \end{aligned}} \tag{17.2.8}$$

Replacing $\boldsymbol{A}, \boldsymbol{A}_m$ in terms of Eq. (17.2.6), we may express the solutions (17.2.7) directly in terms of the current densities:

$$\begin{aligned} \boldsymbol{E} &= \frac{1}{j\omega\epsilon}\int_V \big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla'G - j\omega\epsilon\boldsymbol{J}_m\times\nabla'G\big]dV' \\ \boldsymbol{H} &= \frac{1}{j\omega\mu}\int_V \big[k^2\boldsymbol{J}_m G + (\boldsymbol{J}_m\cdot\nabla')\nabla'G + j\omega\mu\boldsymbol{J}\times\nabla'G\big]dV' \end{aligned} \tag{17.2.9}$$

Alternatively, if we also use the charge densities, we obtain from (17.2.4):

$$\begin{aligned} \boldsymbol{E} &= \int_V \Big[-j\omega\mu\boldsymbol{J}G + \frac{\rho}{\epsilon}\nabla'G - \boldsymbol{J}_m\times\nabla'G\Big]dV' \\ \boldsymbol{H} &= \int_V \Big[-j\omega\epsilon\boldsymbol{J}_m G + \frac{\rho_m}{\mu}\nabla'G + \boldsymbol{J}\times\nabla'G\Big]dV' \end{aligned} \tag{17.2.10}$$

17.3 Radiation Fields from Magnetic Currents

The radiation fields of the solutions (17.2.7) can be obtained by making the far-field approximation, which consists of the replacements:

$$\frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \simeq \frac{e^{-jkr}}{4\pi r}e^{j\boldsymbol{k}\cdot\mathbf{r}'} \quad\text{and}\quad \nabla \simeq -j\boldsymbol{k} \tag{17.3.1}$$

where $\boldsymbol{k} = k\hat{\mathbf{r}}$. Then, the vector potentials of Eq. (17.2.6) take the simplified form:

$$\boldsymbol{A}(\mathbf{r}) = \mu\frac{e^{-jkr}}{4\pi r}\boldsymbol{F}(\theta,\phi), \qquad \boldsymbol{A}_m(\mathbf{r}) = \epsilon\frac{e^{-jkr}}{4\pi r}\boldsymbol{F}_m(\theta,\phi) \tag{17.3.2}$$

where the radiation vectors are the Fourier transforms of the current densities:

$$\boxed{\begin{aligned} \boldsymbol{F}(\theta,\phi) &= \int_V \boldsymbol{J}(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,dV' \\ \boldsymbol{F}_m(\theta,\phi) &= \int_V \boldsymbol{J}_m(\mathbf{r}')e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,dV' \end{aligned}} \qquad \text{(radiation vectors)} \tag{17.3.3}$$

Setting $\boldsymbol{J} = \boldsymbol{J}_m = 0$ in Eq. (17.2.8) because we are evaluating the fields far from the current sources, and using the approximation $\nabla = -j\boldsymbol{k} = -jk\hat{\mathbf{r}}$, and the relationship $k/\epsilon = \omega\eta$, we find the radiated $\boldsymbol{E}$ and $\boldsymbol{H}$ fields:

$$\boxed{\begin{aligned} \boldsymbol{E} &= -j\omega\big[\hat{\mathbf{r}}\times(\boldsymbol{A}\times\hat{\mathbf{r}}) - \eta\,\hat{\mathbf{r}}\times\boldsymbol{A}_m\big] = -jk\frac{e^{-jkr}}{4\pi r}\hat{\mathbf{r}}\times\big[\eta\boldsymbol{F}\times\hat{\mathbf{r}} - \boldsymbol{F}_m\big] \\ \boldsymbol{H} &= -\frac{j\omega}{\eta}\big[\eta\,\hat{\mathbf{r}}\times(\boldsymbol{A}_m\times\hat{\mathbf{r}}) + \hat{\mathbf{r}}\times\boldsymbol{A}\big] = -\frac{jk}{\eta}\frac{e^{-jkr}}{4\pi r}\hat{\mathbf{r}}\times\big[\eta\boldsymbol{F} + \boldsymbol{F}_m\times\hat{\mathbf{r}}\big] \end{aligned}} \tag{17.3.4}$$

These generalize Eq. (14.10.2) to magnetic currents. As in Eq. (14.10.3), we have:

$$\boldsymbol{H} = \frac{1}{\eta}\hat{\mathbf{r}}\times\boldsymbol{E} \tag{17.3.5}$$

Noting that $\hat{\mathbf{r}}\times(\mathbf{F}\times\hat{\mathbf{r}})=\hat{\boldsymbol{\theta}}F_\theta+\hat{\boldsymbol{\phi}}F_\phi$ and $\hat{\mathbf{r}}\times\mathbf{F}=\hat{\boldsymbol{\phi}}F_\theta-\hat{\boldsymbol{\theta}}F_\phi$, and similarly for $\mathbf{F}_m$, we find for the polar components of Eq. (17.3.4):

$$\begin{aligned}\mathbf{E}&=-jk\,\frac{e^{-jkr}}{4\pi r}\big[\hat{\boldsymbol{\theta}}(\eta F_\theta+F_{m\phi})+\hat{\boldsymbol{\phi}}(\eta F_\phi-F_{m\theta})\big]\\ \mathbf{H}&=-\frac{jk}{\eta}\,\frac{e^{-jkr}}{4\pi r}\big[-\hat{\boldsymbol{\theta}}(\eta F_\phi-F_{m\theta})+\hat{\boldsymbol{\phi}}(\eta F_\theta+F_{m\phi})\big]\end{aligned}\tag{17.3.6}$$

The Poynting vector is given by the generalization of Eq. (15.1.1):

$$\boldsymbol{\mathcal{P}}=\frac{1}{2}\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*)=\hat{\mathbf{r}}\,\frac{k^2}{32\pi^2\eta r^2}\big[|\eta F_\theta+F_{m\phi}|^2+|\eta F_\phi-F_{m\theta}|^2\big]=\hat{\mathbf{r}}\,\mathcal{P}_r\tag{17.3.7}$$

and the radiation intensity:

$$U(\theta,\phi)=\frac{dP}{d\Omega}=r^2\mathcal{P}_r=\frac{k^2}{32\pi^2\eta}\big[|\eta F_\theta+F_{m\phi}|^2+|\eta F_\phi-F_{m\theta}|^2\big]\tag{17.3.8}$$

17.4 Radiation Fields from Apertures

For an aperture antenna with effective surface currents given by Eq. (17.1.1), the volume integrations in Eq. (17.2.9) reduce to surface integrations over the aperture A:

$$\begin{aligned}\mathbf{E}&=\frac{1}{j\omega\epsilon}\int_A\big[(\mathbf{J}_s\cdot\nabla')\nabla'G+k^2\mathbf{J}_s\,G-j\omega\epsilon\,\mathbf{J}_{ms}\times\nabla'G\big]dS'\\ \mathbf{H}&=\frac{1}{j\omega\mu}\int_A\big[(\mathbf{J}_{ms}\cdot\nabla')\nabla'G+k^2\mathbf{J}_{ms}\,G+j\omega\mu\,\mathbf{J}_s\times\nabla'G\big]dS'\end{aligned}\tag{17.4.1}$$

and, explicitly in terms of the aperture fields shown in Fig. 17.1.1:

$$\boxed{\begin{aligned}\mathbf{E}&=\frac{1}{j\omega\epsilon}\int_A\big[(\hat{\mathbf{n}}\times\mathbf{H}_a)\cdot\nabla'(\nabla'G)+k^2(\hat{\mathbf{n}}\times\mathbf{H}_a)G+j\omega\epsilon(\hat{\mathbf{n}}\times\mathbf{E}_a)\times\nabla'G\big]dS'\\ \mathbf{H}&=\frac{1}{j\omega\mu}\int_A\big[-(\hat{\mathbf{n}}\times\mathbf{E}_a)\cdot\nabla'(\nabla'G)-k^2(\hat{\mathbf{n}}\times\mathbf{E}_a)G+j\omega\mu(\hat{\mathbf{n}}\times\mathbf{H}_a)\times\nabla'G\big]dS'\end{aligned}}\tag{17.4.2}$$

These are known as *Kottler's formulas* [1116–1121,1111,1122–1126]. We derive them in Sec. 17.12. The equation for $\mathbf{H}$ can also be obtained from that of $\mathbf{E}$ by the application of a duality transformation, that is, $\mathbf{E}_a\to\mathbf{H}_a$, $\mathbf{H}_a\to-\mathbf{E}_a$ and $\epsilon\to\mu$, $\mu\to\epsilon$.

In the far-field limit, the radiation fields are still given by Eq. (17.3.6), but now the radiation vectors are given by the two-dimensional Fourier transform-like integrals over the aperture:

$$\boxed{\begin{aligned}\mathbf{F}(\theta,\phi)&=\int_A\mathbf{J}_s(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dS'=\int_A\hat{\mathbf{n}}\times\mathbf{H}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dS'\\ \mathbf{F}_m(\theta,\phi)&=\int_A\mathbf{J}_{ms}(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dS'=-\int_A\hat{\mathbf{n}}\times\mathbf{E}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dS'\end{aligned}}\tag{17.4.3}$$

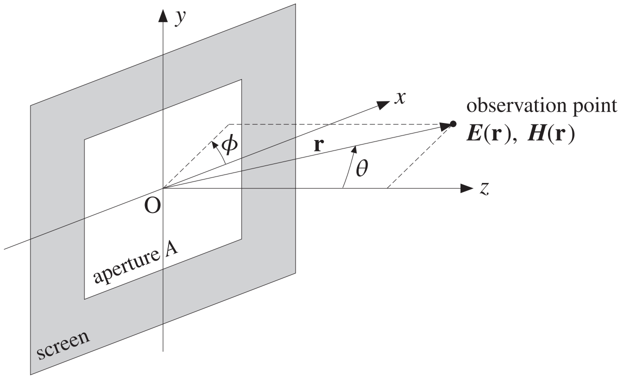


Fig. 17.4.1 Radiation fields from an aperture.

Fig. 17.4.1 shows the polar angle conventions, where we took the origin to be somewhere in the middle of the aperture A.

The aperture surface A and the screen in Fig. 17.1.1 can be arbitrarily curved. However, a common case is to assume that they are both flat. Then, Eqs. (17.4.3) become ordinary 2-d Fourier transform integrals. Taking the aperture plane to be the xy-plane as in Fig. 17.1.1, the aperture normal becomes $\hat{\mathbf{n}}=\hat{\mathbf{z}}$, and thus, it can be taken out of the integrands. Setting $dS'=dx'dy'$, we rewrite Eq. (17.4.3) in the form:

$$\begin{aligned}\mathbf{F}(\theta,\phi)&=\int_A\mathbf{J}_s(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'=\hat{\mathbf{z}}\times\int_A\mathbf{H}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'\\ \mathbf{F}_m(\theta,\phi)&=\int_A\mathbf{J}_{ms}(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'=-\hat{\mathbf{z}}\times\int_A\mathbf{E}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'\end{aligned}\tag{17.4.4}$$

where $e^{j\mathbf{k}\cdot\mathbf{r}'}=e^{jk_xx'+jk_yy'}$ and $k_x=k\cos\phi\sin\theta$, $k_y=k\sin\phi\sin\theta$. It proves convenient then to introduce the *two-dimensional Fourier transforms* of the aperture fields:

$$\boxed{\begin{aligned}\mathbf{f}(\theta,\phi)&=\int_A\mathbf{E}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'=\int_A\mathbf{E}_a(x',y')e^{jk_xx'+jk_yy'}\,dx'dy'\\ \mathbf{g}(\theta,\phi)&=\int_A\mathbf{H}_a(\mathbf{r}')e^{j\mathbf{k}\cdot\mathbf{r}'}\,dx'dy'=\int_A\mathbf{H}_a(x',y')e^{jk_xx'+jk_yy'}\,dx'dy'\end{aligned}}\tag{17.4.5}$$

Then, the radiation vectors become:

$$\begin{aligned}\mathbf{F}(\theta,\phi)&=\hat{\mathbf{z}}\times\mathbf{g}(\theta,\phi)\\ \mathbf{F}_m(\theta,\phi)&=-\hat{\mathbf{z}}\times\mathbf{f}(\theta,\phi)\end{aligned}\tag{17.4.6}$$

Because $\mathbf{E}_a,\mathbf{H}_a$ are tangential to the aperture plane, they can be resolved into their cartesian components, for example, $\mathbf{E}_a=\hat{\mathbf{x}}\,E_{ax}+\hat{\mathbf{y}}\,E_{ay}$. Then, the quantities $\mathbf{f},\mathbf{g}$ can be resolved in the same way, for example, $\mathbf{f}=\hat{\mathbf{x}}\,f_x+\hat{\mathbf{y}}\,f_y$. Thus, we have:

$$\begin{aligned} \boldsymbol{F} &= \hat{\mathbf{z}}\times\boldsymbol{g} = \hat{\mathbf{z}}\times(\hat{\mathbf{x}}g_x+\hat{\mathbf{y}}g_y) = \hat{\mathbf{y}}g_x - \hat{\mathbf{x}}g_y \\ \boldsymbol{F}_m &= -\hat{\mathbf{z}}\times\boldsymbol{f} = -\hat{\mathbf{z}}\times(\hat{\mathbf{x}}f_x+\hat{\mathbf{y}}f_y) = \hat{\mathbf{x}}f_y - \hat{\mathbf{y}}f_x \end{aligned} \tag{17.4.7}$$

The polar components of the radiation vectors are determined as follows:

$$F_\theta = \hat{\boldsymbol{\theta}}\cdot\boldsymbol{F} = \hat{\boldsymbol{\theta}}\cdot(\hat{\mathbf{y}}g_x - \hat{\mathbf{x}}g_y) = g_x\sin\phi\cos\theta - g_y\cos\phi\cos\theta$$

where we read off the dot products $(\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{x}})$ and $(\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{y}})$ from Eq. (14.8.3). The remaining polar components are found similarly, and we summarize them below:

$$\begin{aligned} F_\theta &= -\cos\theta\,(g_y\cos\phi - g_x\sin\phi) \\ F_\phi &= g_x\cos\phi + g_y\sin\phi \\ F_{m\theta} &= \cos\theta\,(f_y\cos\phi - f_x\sin\phi) \\ F_{m\phi} &= -(f_x\cos\phi + f_y\sin\phi) \end{aligned} \tag{17.4.8}$$

It follows from Eq. (17.3.6) that the radiated E-field will be:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{4\pi r}\big[(f_x\cos\phi + f_y\sin\phi) + \eta\cos\theta\,(g_y\cos\phi - g_x\sin\phi)\big] \\ E_\phi &= jk\,\frac{e^{-jkr}}{4\pi r}\big[\cos\theta\,(f_y\cos\phi - f_x\sin\phi) - \eta\,(g_x\cos\phi + g_y\sin\phi)\big] \end{aligned}} \tag{17.4.9}$$

The radiation fields resulting from the alternative forms of the field equivalence principle, Eqs. (17.1.2) and (17.1.3), are obtained from Eq. (17.4.9) by removing the g- or the f-terms and doubling the remaining term. We have for the PEC case:

$$\boxed{\begin{aligned} E_\theta &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[f_x\cos\phi + f_y\sin\phi\big] \\ E_\phi &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[\cos\theta\,(f_y\cos\phi - f_x\sin\phi)\big] \end{aligned}} \tag{17.4.10}$$

and for the PMC case:

$$\boxed{\begin{aligned} E_\theta &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[\eta\cos\theta\,(g_y\cos\phi - g_x\sin\phi)\big] \\ E_\phi &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[-\eta\,(g_x\cos\phi + g_y\sin\phi)\big] \end{aligned}} \tag{17.4.11}$$

In all three cases, the radiated magnetic fields are obtained from:

$$\boxed{H_\theta = -\frac{1}{\eta}E_\phi\,, \qquad H_\phi = \frac{1}{\eta}E_\theta} \tag{17.4.12}$$

We note that Eq. (17.4.9) is the average of Eqs. (17.4.10) and (17.4.11). Also, Eq. (17.4.11) is the dual of Eq. (17.4.10). Indeed, using Eq. (17.4.12), we obtain the following H-components for Eq. (17.4.11), which can be derived from Eq. (17.4.10) by the duality transformation $\boldsymbol{E}_a\to\boldsymbol{H}_a$ or $\boldsymbol{f}\to\boldsymbol{g}$:

$$\begin{aligned} H_\theta &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[g_x\cos\phi + g_y\sin\phi\big] \\ H_\phi &= 2jk\,\frac{e^{-jkr}}{4\pi r}\big[\cos\theta\,(g_y\cos\phi - g_x\sin\phi)\big] \end{aligned} \tag{17.4.13}$$

At $\theta = 90^{\mathrm{o}}$, the components E_ϕ, H_ϕ become tangential to the aperture screen. We note that because of the $\cos\theta$ factors, E_ϕ (resp. H_ϕ) will vanish in the PEC (resp. PMC) case, in accordance with the boundary conditions.

17.5 Huygens Source

The aperture fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ are referred to as Huygens source if at all points on the aperture they are related by the uniform plane-wave relationship:

$$\boxed{\boldsymbol{H}_a = \frac{1}{\eta}\,\hat{\mathbf{n}}\times\boldsymbol{E}_a} \qquad \text{(Huygens source)} \tag{17.5.1}$$

where η is the characteristic impedance of vacuum.

For example, this is the case if a uniform plane wave is incident normally on the aperture plane from the left, as shown in Fig. 17.5.1. The aperture fields are assumed to be equal to the incident fields, $\boldsymbol{E}_a = \boldsymbol{E}_{\mathrm{inc}}$ and $\boldsymbol{H}_a = \boldsymbol{H}_{\mathrm{inc}}$, and the incident fields satisfy $\boldsymbol{H}_{\mathrm{inc}} = \hat{\mathbf{z}}\times\boldsymbol{E}_{\mathrm{inc}}/\eta$.

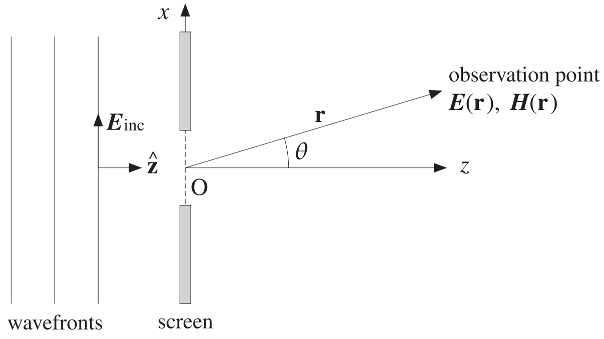

x
observation point
$\boldsymbol{E}(\mathbf{r}),\ \boldsymbol{H}(\mathbf{r})$
r
$\boldsymbol{E}_{\mathrm{inc}}$
θ
$\hat{\mathbf{z}}$
z
O
wavefronts
screen

Fig. 17.5.1 Uniform plane wave incident on an aperture.

The Huygens source condition is not always satisfied. For example, if the uniform plane wave is incident obliquely on the aperture, then η must be replaced by the transverse impedance η_T, which depends on the angle of incidence and the polarization of the incident wave as discussed in Sec. 7.2.

Similarly, if the aperture is the open end of a waveguide, then η must be replaced by the waveguide's transverse impedance, such as η_{TE} or η_{TM}, depending on the assumed waveguide mode. On the other hand, if the waveguide ends are flared out into a horn with a large aperture, then Eq. (17.5.1) is approximately valid.

The Huygens source condition implies the same relationship for the Fourier transforms of the aperture fields, that is, (with $\hat{\mathbf{n}} = \hat{\mathbf{z}}$)

$$\boxed{\mathbf{g} = \frac{1}{\eta}\,\hat{\mathbf{n}}\times\mathbf{f}} \quad\Rightarrow\quad \boxed{g_x = -\frac{1}{\eta}f_y\,,\quad g_y = \frac{1}{\eta}f_x} \tag{17.5.2}$$

Inserting these into Eq. (17.4.9) we may express the radiated electric field in terms of $\mathbf{f}$ only. We find:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,[f_y\cos\phi - f_x\sin\phi] \end{aligned}} \tag{17.5.3}$$

The factor $(1+\cos\theta)/2$ is known as an *obliquity factor*. The PEC case of Eq. (17.4.10) remains unchanged for a Huygens source, but the PMC case becomes:

$$\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,\cos\theta\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,[f_y\cos\phi - f_x\sin\phi] \end{aligned} \tag{17.5.4}$$

We may summarize all three cases by the single formula:

$$\boxed{\begin{aligned} E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,c_\theta\,[f_x\cos\phi + f_y\sin\phi] \\ E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,c_\phi\,[f_y\cos\phi - f_x\sin\phi] \end{aligned}} \quad \text{(fields from Huygens source)} \tag{17.5.5}$$

where the obliquity factors are defined in the three cases:

$$\begin{bmatrix} c_\theta \\ c_\phi \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1+\cos\theta \\ 1+\cos\theta \end{bmatrix},\quad \begin{bmatrix} 1 \\ \cos\theta \end{bmatrix},\quad \begin{bmatrix} \cos\theta \\ 1 \end{bmatrix} \quad \text{(obliquity factors)} \tag{17.5.6}$$

We note that the first is the average of the last two. The obliquity factors are equal to unity in the forward direction $\theta = 0^{\mathrm{o}}$ and vary little for near-forward angles. Therefore, the radiation patterns predicted by the three methods are very similar in their mainlobe behavior.

In the case of a *modified Huygens source* that replaces η by η_T, Eqs. (17.5.5) retain their form. The aperture fields and their Fourier transforms are now assumed to be related by:

$$\boxed{\mathbf{H}_a = \frac{1}{\eta_T}\,\hat{\mathbf{z}}\times\mathbf{E}_a} \quad\Rightarrow\quad \boxed{\mathbf{g} = \frac{1}{\eta_T}\,\hat{\mathbf{z}}\times\mathbf{f}} \tag{17.5.7}$$

Inserting these into Eq. (17.4.9), we obtain the *modified obliquity factors*:

$$c_\theta = \frac{1}{2}\,[1 + K\cos\theta]\,,\quad c_\phi = \frac{1}{2}\,[K + \cos\theta]\,,\qquad K = \frac{\eta}{\eta_T} \tag{17.5.8}$$

17.6 Directivity and Effective Area of Apertures

For any aperture, given the radiation fields E_θ, E_ϕ of Eqs. (17.4.9)–(17.4.11), the corresponding radiation intensity is:

$$U(\theta,\phi) = \frac{dP}{d\Omega} = r^2\mathcal{P}_r = r^2\,\frac{1}{2\eta}\big[|E_\theta|^2 + |E_\phi|^2\big] = r^2\,\frac{1}{2\eta}\,|\mathbf{E}(\theta,\phi)|^2 \tag{17.6.1}$$

Because the aperture radiates only into the right half-space $0 \le \theta \le \pi/2$, the total radiated power and the effective isotropic radiation intensity will be:

$$P_{\mathrm{rad}} = \int_0^{\pi/2}\!\!\int_0^{2\pi} U(\theta,\phi)\,d\Omega\,,\qquad U_I = \frac{P_{\mathrm{rad}}}{4\pi} \tag{17.6.2}$$

The directive gain is computed by $D(\theta,\phi) = U(\theta,\phi)/U_I$, and the normalized gain by $g(\theta,\phi) = U(\theta,\phi)/U_{\mathrm{max}}$. For a typical aperture, the maximum intensity U_{max} is towards the forward direction $\theta = 0^{\mathrm{o}}$. In the case of a Huygens source, we have:

$$U(\theta,\phi) = \frac{k^2}{8\pi^2\eta}\,\big[c_\theta^2|f_x\cos\phi + f_y\sin\phi|^2 + c_\phi^2|f_y\cos\phi - f_x\sin\phi|^2\big] \tag{17.6.3}$$

Assuming that the maximum is towards $\theta = 0^{\mathrm{o}}$, then $c_\theta = c_\phi = 1$, and we find for the maximum intensity:

$$\begin{aligned} U_{\mathrm{max}} &= \frac{k^2}{8\pi^2\eta}\,\big[|f_x\cos\phi + f_y\sin\phi|^2 + |f_y\cos\phi - f_x\sin\phi|^2\big]_{\theta=0} \\ &= \frac{k^2}{8\pi^2\eta}\,\big[|f_x|^2 + |f_y|^2\big]_{\theta=0} = \frac{k^2}{8\pi^2\eta}\,|\mathbf{f}|^2_{\mathrm{max}} \end{aligned}$$

where $|\mathbf{f}|^2_{\mathrm{max}} = \big[|f_x|^2 + |f_y|^2\big]_{\theta=0}$. Setting $k = 2\pi/\lambda$, we have:

$$U_{\mathrm{max}} = \frac{1}{2\lambda^2\eta}\,|\mathbf{f}|^2_{\mathrm{max}} \tag{17.6.4}$$

It follows that the *normalized gain* will be:

$$g(\theta,\phi) = \frac{c_\theta^2|f_x\cos\phi + f_y\sin\phi|^2 + c_\phi^2|f_y\cos\phi - f_x\sin\phi|^2}{|\mathbf{f}|^2_{\mathrm{max}}} \tag{17.6.5}$$

In the case of Eq. (17.4.9) with $c_\theta = c_\phi = (1+\cos\theta)/2$, this simplifies further into:

$$\boxed{g(\theta,\phi) = c_\theta^2\,\frac{|f_x|^2 + |f_y|^2}{|\mathbf{f}|^2_{\mathrm{max}}} = \left(\frac{1+\cos\theta}{2}\right)^2\frac{|\mathbf{f}(\theta,\phi)|^2}{|\mathbf{f}|^2_{\mathrm{max}}}} \tag{17.6.6}$$

The square root of the gain is the (normalized) field strength:

$$\boxed{\frac{|\mathbf{E}(\theta,\phi)|}{|\mathbf{E}|_{\mathrm{max}}} = \sqrt{g(\theta,\phi)} = \left(\frac{1+\cos\theta}{2}\right)\frac{|\mathbf{f}(\theta,\phi)|}{|\mathbf{f}|_{\mathrm{max}}}} \tag{17.6.7}$$

The power computed by Eq. (17.6.2) is the total power that is radiated outwards from a half-sphere of large radius r. An alternative way to compute P_{rad} is to invoke energy conservation and compute the total power that flows *into* the right half-space through the aperture. Assuming a Huygens source, we have:

$$P_{\rm rad} = \int_A \mathcal{P}_z\, dS' = \frac{1}{2}\int_A \hat{\mathbf{z}}\cdot {\rm Re}\big[\boldsymbol{E}_a\times \boldsymbol{H}_a^*\big]\, dS' = \frac{1}{2\eta}\int_A |\boldsymbol{E}_a(\mathbf{r}')|^2 dS' \tag{17.6.8}$$

Because $\theta = 0$ corresponds to $k_x = k_y = 0$, it follows from the Fourier transform definition (17.4.5) that:

$$|\boldsymbol{f}|^2_{\rm max} = \left|\int_A \boldsymbol{E}_a(\mathbf{r}')\, e^{j\mathbf{k}\cdot\mathbf{r}'}\, dS'\right|^2_{k_x=k_y=0} = \left|\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2$$

Therefore, the maximum intensity is given by:

$$U_{\rm max} = \frac{1}{2\lambda^2\eta}\, |\boldsymbol{f}|^2_{\rm max} = \frac{1}{2\lambda^2\eta}\left|\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2 \tag{17.6.9}$$

Dividing (17.6.9) by (17.6.8), we find the directivity:

$$\boxed{D_{\rm max} = 4\pi\,\frac{U_{\rm max}}{P_{\rm rad}} = \frac{4\pi}{\lambda^2}\,\frac{\left|\displaystyle\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2}{\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|^2 dS'} = \frac{4\pi A_{\rm eff}}{\lambda^2}\qquad \text{(directivity)}} \tag{17.6.10}$$

It follows that the maximum effective area of the aperture is:

$$\boxed{A_{\rm eff} = \frac{\left|\displaystyle\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2}{\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|^2 dS'} \le A}\qquad \text{(effective area)} \tag{17.6.11}$$

and the *aperture efficiency*:

$$\boxed{e_a = \frac{A_{\rm eff}}{A} = \frac{\left|\displaystyle\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2}{A\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|^2 dS'} \le 1}\qquad \text{(aperture efficiency)} \tag{17.6.12}$$

The inequalities in Eqs. (17.6.11) and (17.6.12) can be thought of as special cases of the Cauchy-Schwarz inequality. It follows that equality is reached whenever $\boldsymbol{E}_a(\mathbf{r}')$ is uniform over the aperture, that is, independent of $\mathbf{r}'$.

Thus, *uniform* apertures achieve the *highest* directivity and have effective areas *equal* to their geometrical areas.

Because the integrand in the numerator of e_a depends both on the magnitude and the phase of $\boldsymbol{E}_a$, it proves convenient to separate out these effects by defining the *aperture taper* efficiency or *loss*, $e_{\rm atl}$, and the *phase error* efficiency or *loss*, $e_{\rm pel}$, as follows:

$$e_{\rm atl} = \frac{\left|\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|\, dS'\right|^2}{A\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|^2 dS'}\,,\qquad e_{\rm pel} = \frac{\left|\displaystyle\int_A \boldsymbol{E}_a(\mathbf{r}')\, dS'\right|^2}{\left|\displaystyle\int_A |\boldsymbol{E}_a(\mathbf{r}')|\, dS'\right|^2} \tag{17.6.13}$$

so that e_a becomes the product:

$$\boxed{e_a = e_{\rm atl}\, e_{\rm pel}} \tag{17.6.14}$$

17.7 Uniform Apertures

In uniform apertures, the fields $\boldsymbol{E}_a, \boldsymbol{H}_a$ are assumed to be constant over the aperture area. Fig. 17.7.1 shows the examples of a rectangular and a circular aperture. For convenience, we will assume a Huygens source.

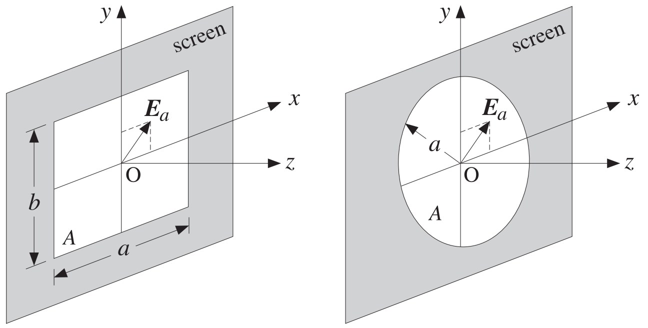


Fig. 17.7.1 Uniform rectangular and circular apertures.

The field $\boldsymbol{E}_a$ can have an arbitrary direction, with constant x- and y-components, $\boldsymbol{E}_a = \hat{\mathbf{x}}\,E_{0x} + \hat{\mathbf{y}}\,E_{0y}$. Because $\boldsymbol{E}_a$ is constant, its Fourier transform $\boldsymbol{f}(\theta,\phi)$ becomes:

$$\boldsymbol{f}(\theta,\phi) = \int_A \boldsymbol{E}_a(\mathbf{r}')\, e^{j\mathbf{k}\cdot\mathbf{r}'}\, dS' = \boldsymbol{E}_a\int_A e^{j\mathbf{k}\cdot\mathbf{r}'}\, dS' \equiv A\, f(\theta,\phi)\, \boldsymbol{E}_a \tag{17.7.1}$$

where we introduced the normalized scalar quantity:

$$\boxed{f(\theta,\phi) = \frac{1}{A}\int_A e^{j\mathbf{k}\cdot\mathbf{r}'}\, dS'}\qquad \text{(uniform-aperture pattern)} \tag{17.7.2}$$

The quantity $f(\theta,\phi)$ depends on the assumed geometry of the aperture and it, alone, determines the radiation pattern. Noting that the quantity $|\boldsymbol{E}_a|$ cancels out from the ratio in the gain (17.6.7) and that $f(0,\phi) = (1/A)\int_A dS' = 1$, we find for the normalized gain and field strengths:

$$\boxed{\frac{|\boldsymbol{E}(\theta,\phi)|}{|\boldsymbol{E}|_{\rm max}} = \sqrt{g(\theta,\phi)} = \left(\frac{1+\cos\theta}{2}\right)|f(\theta,\phi)|} \tag{17.7.3}$$

17.8 Rectangular Apertures

For a rectangular aperture of sides a, b, the area integral (17.7.2) is separable in the x- and y-directions:

$$f(\theta,\phi) = \frac{1}{ab}\int_{-a/2}^{a/2}\int_{-b/2}^{b/2} e^{jk_xx'+jk_yy'}\, dx'dy' = \frac{1}{a}\int_{-a/2}^{a/2} e^{jk_xx'}\, dx'\cdot\frac{1}{b}\int_{-b/2}^{b/2} e^{jk_yy'}\, dy'$$

where we placed the origin of the $\mathbf{r}'$ integration in the middle of the aperture. The above integrals result in the sinc-function patterns:

$$\boxed{f(\theta,\phi)=\frac{\sin(k_xa/2)}{k_xa/2}\,\frac{\sin(k_yb/2)}{k_yb/2}=\frac{\sin(\pi v_x)}{\pi v_x}\,\frac{\sin(\pi v_y)}{\pi v_y}}\tag{17.8.1}$$

where we defined the quantities v_x, v_y:

$$\begin{aligned} v_x &= \frac{1}{2\pi}\,k_xa = \frac{1}{2\pi}\,ka\sin\theta\cos\phi = \frac{a}{\lambda}\,\sin\theta\cos\phi \\ v_y &= \frac{1}{2\pi}\,k_yb = \frac{1}{2\pi}\,kb\sin\theta\sin\phi = \frac{b}{\lambda}\,\sin\theta\sin\phi \end{aligned}\tag{17.8.2}$$

The pattern simplifies along the two principal planes, the xz- and yz-planes, corresponding to $\phi = 0^{\rm o}$ and $\phi = 90^{\rm o}$. We have:

$$\begin{aligned} f(\theta,0^{\rm o}) &= \frac{\sin(\pi v_x)}{\pi v_x} = \frac{\sin\big((\pi a/\lambda)\sin\theta\big)}{(\pi a/\lambda)\sin\theta} \\ f(\theta,90^{\rm o}) &= \frac{\sin(\pi v_y)}{\pi v_y} = \frac{\sin\big((\pi b/\lambda)\sin\theta\big)}{(\pi b/\lambda)\sin\theta} \end{aligned}\tag{17.8.3}$$

Fig. 17.8.1 shows the three-dimensional pattern of Eq. (17.7.3) as a function of the independent variables v_x, v_y, for aperture dimensions $a = 8\lambda$ and $b = 4\lambda$. The x, y separability of the pattern is evident. The essential MATLAB code for generating this figure was (note MATLAB's definition of $\text{sinc}(x)=\sin(\pi x)/(\pi x)$):

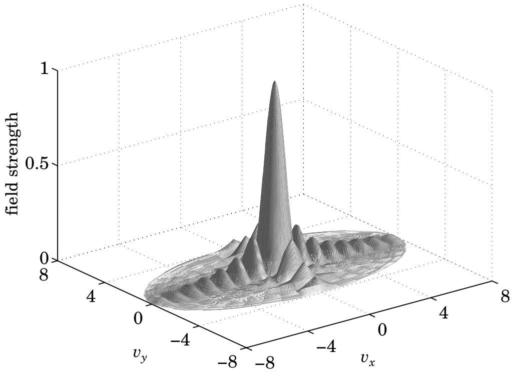


Fig. 17.8.1 Radiation pattern of rectangular aperture ($a = 8\lambda$, $b = 4\lambda$).

```
a = 8; b = 4;
[theta,phi] = meshgrid(0:1:90, 0:9:360);
theta = theta*pi/180;  phi = phi*pi/180;

vx = a*sin(theta).*cos(phi);
vy = b*sin(theta).*sin(phi);

E = abs((1 + cos(theta))/2 .* sinc(vx) .* sinc(vy));
```

```
surfl(vx,vy,E);
shading interp;  colormap(gray(16));
```

As the polar angles vary over $0 \le \theta \le 90^{\rm o}$ and $0 \le \phi \le 360^{\rm o}$, the quantities v_x and v_y vary over the limits $-a/\lambda \le v_x \le a/\lambda$ and $-b/\lambda \le v_y \le b/\lambda$. In fact, the *physically realizable* values of v_x, v_y are those that lie in the ellipse in the v_xv_y-plane:

$$\boxed{\frac{v_x^2}{a^2}+\frac{v_y^2}{b^2}\le\frac{1}{\lambda^2}}\quad\text{(visible region)}\tag{17.8.4}$$

The realizable values of v_x, v_y are referred to as the *visible region.* The graph in Fig. 17.8.1 restricts the values of v_x, v_y within that region.

The radiation pattern consists of a narrow mainlobe directed towards the forward direction $\theta = 0^{\rm o}$ and several sidelobes.

We note the three characteristic properties of the sinc-function patterns: (a) the 3-dB width in v-space is $\Delta v_x = 0.886$ (the 3-dB wavenumber is $v_x = 0.443$); (b) the first sidelobe is down by about 13.26 dB from the mainlobe and occurs at $v_x = 1.4303$; and (c) the first null occurs at $v_x = 1$. See Sec. 19.7 for the proof of these results.

The 3-dB width in angle space can be obtained by linearizing the relationship $v_x = (a/\lambda)\sin\theta$ about $\theta = 0^{\rm o}$, that is, $\Delta v_x = (a/\lambda)\Delta\theta\cos\theta\big|_{\theta=0} = a\Delta\theta/\lambda$. Thus, $\Delta\theta = \lambda\Delta v_x/a$. This ignores also the effect of the obliquity factor. It follows that the 3-dB widths in the two principal planes are (in radians and in degrees):

$$\Delta\theta_x = 0.886\,\frac{\lambda}{a} = 50.76^{\rm o}\,\frac{\lambda}{a}\,,\quad \Delta\theta_y = 0.886\,\frac{\lambda}{b} = 50.76^{\rm o}\,\frac{\lambda}{b}\tag{17.8.5}$$

The 3-dB angles are $\theta_x = \Delta\theta_x/2 = 25.4^{\rm o}\,\lambda/a$ and $\theta_y = \Delta\theta_y/2 = 25.4^{\rm o}\,\lambda/b$. Fig. 17.8.2 shows the two principal radiation patterns of Eq. (17.7.3) as functions of θ, for the case $a = 8\lambda$, $b = 4\lambda$. The obliquity factor was included, but it makes essentially no difference near the mainlobe and first sidelobe region, ultimately suppressing the response at $\theta = 90^{\rm o}$ by a factor of 0.5.

The 3-dB widths are shown on the graphs. The first sidelobes occur at the angles $\theta_a = \text{asin}(1.4303\lambda/a)= 10.30^{\rm o}$ and $\theta_b = \text{asin}(1.4303\lambda/b)= 20.95^{\rm o}$.

For aperture antennas, the gain is approximately equal to the directivity because the losses tend to be very small. The gain of the uniform rectangular aperture is, therefore, $G \simeq D = 4\pi(ab)/\lambda^2$. Multiplying G by Eqs. (17.8.5), we obtain the *gain-beamwidth product* $p = G\,\Delta\theta_x\,\Delta\theta_y = 4\pi(0.886)^2 = 9.8646\ \text{rad}^2 = 32\,383\ \text{deg}^2$. Thus, we have an example of the general formula (15.3.14) (with the angles in radians and in degrees):

$$G = \frac{9.8646}{\Delta\theta_x\,\Delta\theta_y} = \frac{32\,383}{\Delta\theta_x^{\rm o}\,\Delta\theta_y^{\rm o}}\tag{17.8.6}$$

17.9 Circular Apertures

For a circular aperture of radius a, the pattern integral (17.7.2) can be done conveniently using cylindrical coordinates. The cylindrical symmetry implies that $f(\theta,\phi)$ will be independent of ϕ.

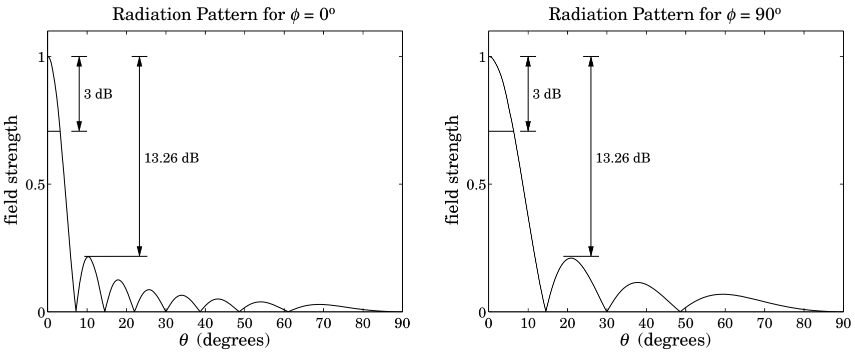


Fig. 17.8.2 Radiation patterns along the two principal planes ($a = 8\lambda$, $b = 4\lambda$).

Therefore, for the purpose of computing the integral (17.7.2), we may set $\phi = 0$. We have then $\boldsymbol{k}\cdot\mathbf{r}' = k_x x' = k\rho'\sin\theta\cos\phi'$. Writing $dS' = \rho' d\rho' d\phi'$, we have:

$$f(\theta) = \frac{1}{\pi a^2}\int_0^a\int_0^{2\pi} e^{jk\rho'\sin\theta\cos\phi'}\rho'\,d\rho' d\phi' \tag{17.9.1}$$

The ϕ'- and ρ'-integrations can be done using the following integral representations for the Bessel functions $J_0(x)$ and $J_1(x)$ [1267]:

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{jx\cos\phi'}\,d\phi' \quad \text{and} \quad \int_0^1 J_0(xr)\,r\,dr = \frac{J_1(x)}{x} \tag{17.9.2}$$

Then Eq. (17.9.1) gives:

$$\boxed{f(\theta) = 2\,\frac{J_1(ka\sin\theta)}{ka\sin\theta} = 2\,\frac{J_1(2\pi u)}{2\pi u}}\,, \qquad u = \frac{1}{2\pi}\,ka\sin\theta = \frac{a}{\lambda}\,\sin\theta \tag{17.9.3}$$

This is the well-known Airy pattern [598] for a circular aperture. The function $f(\theta)$ is normalized to unity at $\theta = 0^o$, because $J_1(x)$ behaves like $J_1(x) \simeq x/2$ for small x.

Fig. 17.9.1 shows the three-dimensional field pattern (17.7.3) as a function of the independent variables $v_x = (a/\lambda)\sin\theta\cos\phi$ and $v_y = (a/\lambda)\sin\theta\sin\phi$, for an aperture radius of $a = 3\lambda$. The obliquity factor was not included as it makes little difference near the main lobe. The MATLAB code for this graph was implemented with the built-in function `besselj`:

```
a = 3;
[theta,phi] = meshgrid(0:1:90, 0:9:360);
theta = theta*pi/180;  phi = phi*pi/180;

vx = a*sin(theta).*cos(phi);
vy = a*sin(theta).*sin(phi);
u = a*sin(theta);

E = ones(size(u));
i = find(u);
```

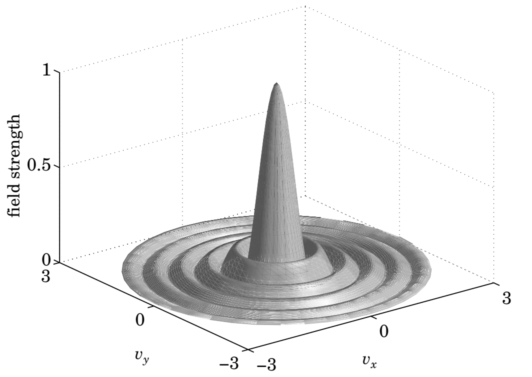


Fig. 17.9.1 Radiation pattern of circular aperture ($a = 3\lambda$).

```
E(i) = abs(2*besselj(1,2*pi*u(i))./(2*pi*u(i)));

surfl(vx,vy,E);
shading interp;  colormap(gray(16));
```

The visible region is the circle on the $v_x v_y$-plane:

$$v_x^2 + v_y^2 \le \frac{a^2}{\lambda^2} \tag{17.9.4}$$

The mainlobe/sidelobe characteristics of $f(\theta)$ are as follows. The 3-dB wavenumber is $u = 0.2572$ and the 3-dB width in u-space is $\Delta u = 2\times 0.2572 = 0.5144$. The first null occurs at $u = 0.6098$ so that the first-null width is $\Delta u = 2\times 0.6098 = 1.22$. The first sidelobe occurs at $u = 0.8174$ and its height is $|f(u)| = 0.1323$ or 17.56 dB below the mainlobe. The beamwidths in angle space can be obtained from $\Delta u = a(\Delta\theta)/\lambda$, which gives for the 3-dB and first-null widths in radians and degrees:

$$\Delta\theta_{\text{3dB}} = 0.5144\,\frac{\lambda}{a} = 29.47^o\,\frac{\lambda}{a}\,, \qquad \Delta\theta_{\text{null}} = 1.22\,\frac{\lambda}{a} = 70^o\,\frac{\lambda}{a} \tag{17.9.5}$$

The 3-dB angle is $\theta_{\text{3dB}} = \Delta\theta_{\text{3dB}}/2 = 0.2572\lambda/a = 14.74^o\,\lambda/a$ and the first-null angle $\theta_{\text{null}} = 0.6098\lambda/a$. Fig. 17.9.2 shows the radiation pattern of Eq. (17.7.3) as a function of θ, for the case $a = 3\lambda$. The obliquity factor was included.

The graph shows the 3-dB width and the first sidelobe, which occurs at the angle $\theta_a = \text{asin}(0.817\lambda/a) = 15.8^o$. The first null occurs at $\theta_{\text{null}} = \text{asin}(0.6098\lambda/a) = 11.73^o$, whereas the approximation $\theta_{\text{null}} = 0.6098\lambda/a$ gives 11.65^o.

The gain-beamwidth product is $p = G(\Delta\theta_{\text{3dB}})^2 = \big[4\pi(\pi a^2)/\lambda^2\big](0.514\lambda/a)^2 = 4\pi^2(0.5144)^2 = 10.4463\text{ rad}^2 = 34\,293\text{ deg}^2$. Thus, in radians and degrees:

$$G = \frac{10.4463}{(\Delta\theta_{\text{3dB}})^2} = \frac{34\,293}{(\Delta\theta^o_{\text{3dB}})^2} \tag{17.9.6}$$

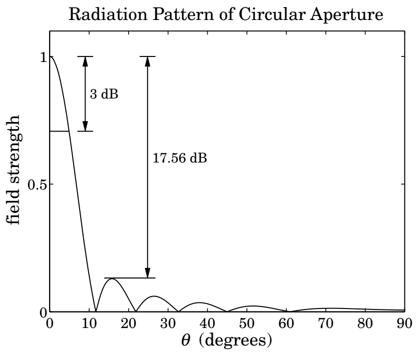


Fig. 17.9.2 Radiation pattern of circular aperture ($a = 3\lambda$).

The first-null angle $\theta_{\text{null}} = 0.6098\lambda/a$ is the so-called *Rayleigh diffraction limit* for the nominal angular resolution of optical instruments, such as microscopes and telescopes. It is usually stated in terms of the diameter $D = 2a$ of the optical aperture:

$$\boxed{\Delta\theta = 1.22\,\frac{\lambda}{D} = 70^{\circ}\,\frac{\lambda}{D}} \qquad \text{(Rayleigh limit)} \tag{17.9.7}$$

17.10 Vector Diffraction Theory

In this section, we provide a justification of the field equivalence principle (17.1.1) and Kottler's formulas (17.4.2) from the point of view of vector diffraction theory. We also discuss the Stratton-Chu and Franz formulas. A historical overview of this subject is given in [1125,1126].

In Sec. 17.2, we worked with the vector potentials and derived the fields due to electric and magnetic currents radiating in an unbounded region. Here, we consider the problem of finding the fields in a volume V bounded by a closed surface S and an infinite spherical surface S_∞, as shown in Fig. 17.10.1.

The solution of this problem requires that we know the current sources within V *and* the electric and magnetic fields tangential to the surface S. The fields $\boldsymbol{E}_1, \boldsymbol{H}_1$ and current sources inside the volume V_1 enclosed by S have an effect on the outside only through the tangential fields on the surface.

We start with Maxwell's equations (17.2.1), which include both electric and magnetic currents. This will help us identify the effective surface currents and derive the field equivalence principle.

Taking the curls of both sides of Ampère's and Faraday's laws and using the vector identity $\nabla\times(\nabla\times\boldsymbol{E}) = \nabla(\nabla\cdot\boldsymbol{E}) - \nabla^2\boldsymbol{E}$, we obtain the following inhomogeneous Helmholtz equations (which are duals of each other):

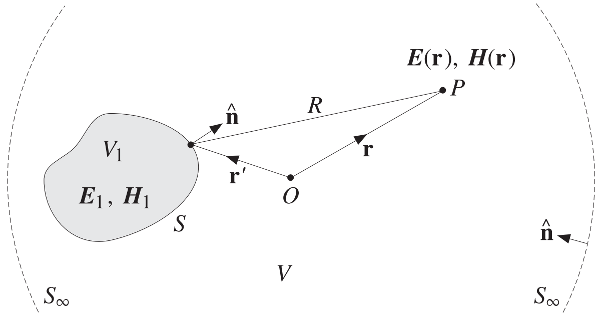


Fig. 17.10.1 Fields outside a closed surface S.

$$\begin{aligned}
\nabla^2\boldsymbol{E} + k^2\boldsymbol{E} &= j\omega\mu\,\boldsymbol{J} + \frac{1}{\epsilon}\nabla\rho + \nabla\times\boldsymbol{J}_m \\
\nabla^2\boldsymbol{H} + k^2\boldsymbol{H} &= j\omega\epsilon\,\boldsymbol{J}_m + \frac{1}{\mu}\nabla\rho_m - \nabla\times\boldsymbol{J}
\end{aligned} \tag{17.10.1}$$

We recall that the Green's function for the Helmholtz equation is:

$$\nabla'^2 G + k^2 G = -\delta^{(3)}(\mathbf{r}-\mathbf{r}')\,, \quad G(\mathbf{r}-\mathbf{r}') = \frac{e^{-jk|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{17.10.2}$$

where $\boldsymbol{\nabla}'$ is the gradient with respect to $\mathbf{r}'$. Applying Green's second identity given by Eq. (C.27) of Appendix C, we obtain:

$$\int_V \big[G\nabla'^2\boldsymbol{E} - \boldsymbol{E}\nabla'^2 G\big]\,dV' = -\oint_{S+S_\infty}\left[G\,\frac{\partial\boldsymbol{E}}{\partial n'} - \boldsymbol{E}\,\frac{\partial G}{\partial n'}\right]dS'\,, \qquad \frac{\partial}{\partial n'} = \hat{\mathbf{n}}\cdot\boldsymbol{\nabla}'$$

where G and $\boldsymbol{E}$ stand for $G(\mathbf{r}-\mathbf{r}')$ and $\boldsymbol{E}(\mathbf{r}')$ and the integration is over $\mathbf{r}'$. The quantity $\partial/\partial n'$ is the directional derivative along $\hat{\mathbf{n}}$. The negative sign in the right-hand side arises from using a unit vector $\hat{\mathbf{n}}$ that is pointing *into* the volume V.

The integral over the infinite surface is taken to be zero. This may be justified more rigorously [1118] by assuming that $\boldsymbol{E}$ and $\boldsymbol{H}$ behave like radiation fields with asymptotic form $E \to \text{const.}\, e^{-jkr}/r$ and $\boldsymbol{H} \to \hat{\mathbf{r}}\times\boldsymbol{E}/\eta$.[†] Thus, dropping the S_∞ term, and adding and subtracting $k^2 G\,\boldsymbol{E}$ in the left-hand side, we obtain:

$$\int_V \big[G(\nabla'^2\boldsymbol{E} + k^2\boldsymbol{E}) - \boldsymbol{E}(\nabla'^2 G + k^2 G)\big]\,dV' = -\oint_S\left[G\,\frac{\partial\boldsymbol{E}}{\partial n'} - \boldsymbol{E}\,\frac{\partial G}{\partial n'}\right]dS' \tag{17.10.3}$$

Using Eq. (17.10.2), the second term on the left may be integrated to give $\boldsymbol{E}(\mathbf{r})$:

$$-\int_V \boldsymbol{E}(\mathbf{r}')\,(\nabla'^2 G + k^2 G)\,dV' = \int_V \boldsymbol{E}(\mathbf{r}')\,\delta^{(3)}(\mathbf{r}-\mathbf{r}')\,dV' = \boldsymbol{E}(\mathbf{r})$$

where we assumed that $\mathbf{r}$ lies in V. This integral is zero if $\mathbf{r}$ lies in V_1 because then $\mathbf{r}'$ can never be equal to $\mathbf{r}$. For arbitrary $\mathbf{r}$, we may write:

[†] The precise conditions are: $r|\boldsymbol{E}| \to$ const. and $r|\boldsymbol{E} - \eta\boldsymbol{H}\times\hat{\mathbf{r}}| \to 0$ as $r \to \infty$.

$$\int_V \boldsymbol{E}(\mathbf{r}')\,\delta^{(3)}(\mathbf{r}-\mathbf{r}')\,dV' = u_V(\mathbf{r})\,\boldsymbol{E}(\mathbf{r}) = \begin{cases} \boldsymbol{E}(\mathbf{r}), & \text{if } \mathbf{r}\in V \\ 0, & \text{if } \mathbf{r}\notin V \end{cases} \tag{17.10.4}$$

where $u_V(\mathbf{r})$ is the characteristic function of the volume region V:[†]

$$u_V(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{r}\in V \\ 0, & \text{if } \mathbf{r}\notin V \end{cases} \tag{17.10.5}$$

We may now solve Eq. (17.10.3) for $\boldsymbol{E}(\mathbf{r})$. In a similar fashion, or, performing a duality transformation on the expression for $\boldsymbol{E}(\mathbf{r})$, we also obtain the corresponding magnetic field $\boldsymbol{H}(\mathbf{r})$. Using (17.10.1), we have:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= \int_V \Big[-j\omega\mu\, G\,\boldsymbol{J} - \frac{1}{\epsilon}G\,\nabla'\rho - G\,\nabla'\times\boldsymbol{J}_m\Big]\,dV' + \oint_S \Big[\boldsymbol{E}\,\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol{E}}{\partial n'}\Big]\,dS' \\ \boldsymbol{H}(\mathbf{r}) &= \int_V \Big[-j\omega\epsilon\, G\,\boldsymbol{J}_m - \frac{1}{\mu}G\,\nabla'\rho_m + G\,\nabla'\times\boldsymbol{J}\Big]\,dV' + \oint_S \Big[\boldsymbol{H}\,\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol{H}}{\partial n'}\Big]\,dS' \end{aligned}} \tag{17.10.6}$$

Because of the presence of the particular surface term, we will refer to these as the *Kirchhoff diffraction formulas.* Eqs. (17.10.6) can be transformed into the so-called *Stratton-Chu formulas* [1116–1121,1111,1122–1126]:[‡]

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) = &\int_V \Big[-j\omega\mu\, G\,\boldsymbol{J} + \frac{\rho}{\epsilon}\,\nabla' G - \boldsymbol{J}_m\times\nabla' G\Big]\,dV' \\ &+ \oint_S \big[-j\omega\mu\, G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\,dS' \\ \boldsymbol{H}(\mathbf{r}) = &\int_V \Big[-j\omega\epsilon\, G\,\boldsymbol{J}_m + \frac{\rho_m}{\mu}\,\nabla' G + \boldsymbol{J}\times\nabla' G\Big]\,dV' \\ &+ \oint_S \big[\,j\omega\epsilon\, G\,(\hat{\mathbf{n}}\times\boldsymbol{E}) + (\hat{\mathbf{n}}\cdot\boldsymbol{H})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\big]\,dS' \end{aligned}} \tag{17.10.7}$$

The proof of the equivalence of (17.10.6) and (17.10.7) is rather involved. Problem 17.4 breaks down the proof into its essential steps.

Term by term comparison of the volume and surface integrals in (17.10.7) yields the effective surface currents of the field equivalence principle:*

$$\boxed{\boldsymbol{J}_s = \hat{\mathbf{n}}\times\boldsymbol{H}\,, \quad \boldsymbol{J}_{ms} = -\hat{\mathbf{n}}\times\boldsymbol{E}} \tag{17.10.8}$$

Similarly, the effective surface charge densities are:

$$\rho_s = \epsilon\,\hat{\mathbf{n}}\cdot\boldsymbol{E}\,, \qquad \rho_{ms} = \mu\,\hat{\mathbf{n}}\cdot\boldsymbol{H} \tag{17.10.9}$$

[†] Technically [1124], one must set $u_V(\mathbf{r}) = 1/2$, if $\mathbf{r}$ lies on the boundary of V, that is, on S.

[‡] See [1113,1119,1125,1126] for earlier work by Larmor, Tedone, Ignatowski, and others.

[*] Initially derived by Larmor and Love [1125,1126], and later developed fully by Schelkunoff [1112,1114].

Eqs. (17.10.7) may be transformed into the *Kottler formulas* [1116–1121,1111,1122–1126], which eliminate the charge densities ρ, ρ_m in favor of the currents $\boldsymbol{J}, \boldsymbol{J}_m$:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) = &\frac{1}{j\omega\epsilon}\int_V \Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\,dV' \\ &+\frac{1}{j\omega\epsilon}\oint_S \big[k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\,dS' \\ \boldsymbol{H}(\mathbf{r}) = &\frac{1}{j\omega\mu}\int_V \Big[k^2\boldsymbol{J}_m G + (\boldsymbol{J}_m\cdot\nabla')\nabla' G + j\omega\mu\,\boldsymbol{J}\times\nabla' G\Big]\,dV' \\ &+\frac{1}{j\omega\mu}\oint_S \big[-k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{E}) - \big((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla'\big)\nabla' G + j\omega\mu\,(\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\big]\,dS' \end{aligned}} \tag{17.10.10}$$

The steps of the proof are outlined in Problem 17.5.

A related problem is to consider a volume V bounded by the surface S, as shown in Fig. 17.10.2. The fields inside V are still given by (17.10.7), with $\hat{\mathbf{n}}$ pointing again into the volume V. If the surface S recedes to infinity, then (17.10.10) reduce to (17.2.9).

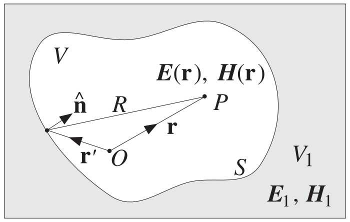


Fig. 17.10.2 Fields inside a closed surface S.

Finally, the Kottler formulas may be transformed into the *Franz formulas* [1121,1111,1122–1124], which are essentially equivalent to Eq. (17.2.8) amended by the vector potentials due to the equivalent surface currents:

$$\boxed{\begin{aligned} \boldsymbol{E}(\mathbf{r}) &= \frac{1}{j\omega\mu\epsilon}\big[\nabla\times\big(\nabla\times(\boldsymbol{A}+\boldsymbol{A}_s)\big) - \mu\,\boldsymbol{J}\big] - \frac{1}{\epsilon}\,\nabla\times(\boldsymbol{A}_m+\boldsymbol{A}_{ms}) \\ \boldsymbol{H}(\mathbf{r}) &= \frac{1}{j\omega\mu\epsilon}\big[\nabla\times\big(\nabla\times(\boldsymbol{A}_m+\boldsymbol{A}_{ms})\big) - \epsilon\,\boldsymbol{J}_m\big] + \frac{1}{\mu}\,\nabla\times(\boldsymbol{A}+\boldsymbol{A}_s) \end{aligned}} \tag{17.10.11}$$

where $\boldsymbol{A}$ and $\boldsymbol{A}_m$ were defined in Eq. (17.2.6). The new potentials are defined by:

$$\begin{aligned} \boldsymbol{A}_s(\mathbf{r}) &= \oint_S \mu\,\boldsymbol{J}_s(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dS' = \oint_S \mu\,\big[\hat{\mathbf{n}}\times\boldsymbol{H}(\mathbf{r}')\big]\,G(\mathbf{r}-\mathbf{r}')\,dS' \\ \boldsymbol{A}_{ms}(\mathbf{r}) &= \oint_S \epsilon\,\boldsymbol{J}_{ms}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dS' = -\oint_S \epsilon\,\big[\hat{\mathbf{n}}\times\boldsymbol{E}(\mathbf{r}')\big]\,G(\mathbf{r}-\mathbf{r}')\,dS' \end{aligned} \tag{17.10.12}$$

Next, we specialize the above formulas to the case where the volume V contains no current sources ($\boldsymbol{J} = \boldsymbol{J}_m = 0$), so that the $\boldsymbol{E}, \boldsymbol{H}$ fields are given only in terms of the surface integral terms.

This happens if we choose S in Fig. 17.10.1 such that all the current sources are inside it, or, if in Fig. 17.10.2 we choose S such that all the current sources are outside it, then, the Kirchhoff, Stratton-Chu, Kottler, and Franz formulas simplify into:

$$\boxed{\begin{aligned}\boldsymbol{E}(\mathbf{r}) &= \oint_S \left[\boldsymbol{E}\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol{E}}{\partial n'}\right] dS' \\ &= \oint_S \big[-j\omega\mu\, G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\, dS' \\ &= \frac{1}{j\omega\epsilon}\oint_S \big[k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\, dS' \\ &= \frac{1}{j\omega\epsilon}\,\nabla\times\Big(\nabla\times\oint_S G\,(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'\Big) + \nabla\times\oint_S G\,(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'\end{aligned}} \tag{17.10.13}$$

$$\boxed{\begin{aligned}\boldsymbol{H}(\mathbf{r}) &= \oint_S \left[\boldsymbol{H}\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol{H}}{\partial n'}\right] dS' \\ &= \oint_S \big[j\omega\epsilon\, G\,(\hat{\mathbf{n}}\times\boldsymbol{E}) + (\hat{\mathbf{n}}\cdot\boldsymbol{H})\,\nabla' G + (\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\big]\, dS' \\ &= \frac{1}{j\omega\mu}\oint_S \big[-k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{E}) - \big((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla'\big)\nabla' G + j\omega\mu\,(\hat{\mathbf{n}}\times\boldsymbol{H})\times\nabla' G\big]\, dS' \\ &= -\frac{1}{j\omega\mu}\,\nabla\times\Big(\nabla\times\oint_S G\,(\hat{\mathbf{n}}\times\boldsymbol{E})\,dS'\Big) + \nabla\times\oint_S G\,(\hat{\mathbf{n}}\times\boldsymbol{H})\,dS'\end{aligned}} \tag{17.10.14}$$

where the last equations are the Franz formulas with $\boldsymbol{A} = \boldsymbol{A}_m = 0$.

Fig. 17.10.3 illustrates the geometry of the two cases. Eqs. (17.10.13) and (17.10.14) represent the vectorial formulation of the Huygens-Fresnel principle, according to which the tangential fields on the surface can be considered to be the sources of the fields away from the surface.

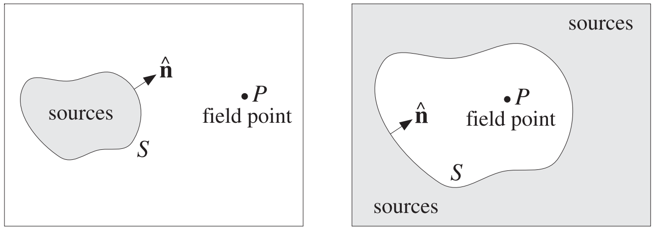


Fig. 17.10.3 Current sources are outside the field region.

17.11 Extinction Theorem

In all of the equivalent formulas for $\boldsymbol{E}(\mathbf{r}), \boldsymbol{H}(\mathbf{r})$, we assumed that $\mathbf{r}$ lies within the volume V. The origin of the left-hand sides in these formulas can be traced to Eq. (17.10.4), and therefore, if $\mathbf{r}$ is not in V but is within the complementary volume V_1, then the left-hand sides of all the formulas are *zero*. This does not mean that the fields inside V_1 are zero—it only means that the sum of the terms on the right-hand sides are zero.

To clarify these remarks, we consider an imaginary closed surface S dividing all space in two volumes V_1 and V, as shown in Fig. 17.11.1. We assume that there are current sources in both regions V and V_1. The surface S_1 is the same as S but its unit vector $\hat{\mathbf{n}}_1$ points into V_1, so that $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}$. Applying (17.10.10) to the volume V, we have:

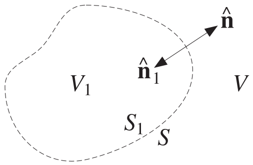


Fig. 17.11.1 Current sources may exist in both V and V_1.

$$\begin{aligned}&\frac{1}{j\omega\epsilon}\oint_S \big[k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\, dS' \\ &+\frac{1}{j\omega\epsilon}\int_V \Big[k^2 \boldsymbol{J} G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}\boldsymbol{E}(\mathbf{r}), & \text{if } \mathbf{r}\in V \\ 0, & \text{if } \mathbf{r}\in V_1\end{cases}\end{aligned}$$

The vanishing of the right-hand side when $\mathbf{r}$ is in V_1 is referred to as an *extinction theorem*.[†] Applying (17.10.10) to V_1, and denoting by $\boldsymbol{E}_1, \boldsymbol{H}_1$ the fields in V_1, we have:

$$\begin{aligned}&\frac{1}{j\omega\epsilon}\oint_{S_1} \big[k^2 G\,(\hat{\mathbf{n}}_1\times\boldsymbol{H}_1) + \big((\hat{\mathbf{n}}_1\times\boldsymbol{H}_1)\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf{n}}_1\times\boldsymbol{E}_1)\times\nabla' G\big]\, dS' \\ &+\frac{1}{j\omega\epsilon}\int_{V_1} \Big[k^2 \boldsymbol{J} G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}0, & \text{if } \mathbf{r}\in V \\ \boldsymbol{E}_1(\mathbf{r}), & \text{if } \mathbf{r}\in V_1\end{cases}\end{aligned}$$

Because $\hat{\mathbf{n}}_1 = -\hat{\mathbf{n}}$, and on the surface $\boldsymbol{E}_1 = \boldsymbol{E}$ and $\boldsymbol{H}_1 = \boldsymbol{H}$, we may rewrite:

$$\begin{aligned}&-\frac{1}{j\omega\epsilon}\oint_S \big[k^2 G\,(\hat{\mathbf{n}}\times\boldsymbol{H}) + \big((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla' G\big]\, dS' \\ &+\frac{1}{j\omega\epsilon}\int_{V_1} \Big[k^2 \boldsymbol{J} G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}0, & \text{if } \mathbf{r}\in V \\ \boldsymbol{E}_1(\mathbf{r}), & \text{if } \mathbf{r}\in V_1\end{cases}\end{aligned}$$

Adding up the two cases and combining the volume integrals into a single one, we obtain:

$$\frac{1}{j\omega\epsilon}\int_{V+V_1}\Big[(\boldsymbol{J}\cdot\nabla')\nabla' G + k^2 G\boldsymbol{J} - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\, dV' = \begin{cases}\boldsymbol{E}(\mathbf{r}), & \text{if } \mathbf{r}\in V \\ \boldsymbol{E}_1(\mathbf{r}), & \text{if } \mathbf{r}\in V_1\end{cases}$$

This is equivalent to Eq. (17.2.9) in which the currents are radiating into unbounded space. We can also see how the sources within V_1 make themselves felt on the outside only through the tangential fields at the surface S, that is, for $\mathbf{r}\in V$:

[†] In fact, it can be used to prove the Ewald-Oseen extinction theorem that we considered in Sec. 14.6.

$$\frac{1}{j\omega\epsilon}\int_{V_1}\Big[k^2\boldsymbol{J}G + (\boldsymbol{J}\cdot\nabla')\nabla' G - j\omega\epsilon\,\boldsymbol{J}_m\times\nabla' G\Big]\,dV'$$
$$=\frac{1}{j\omega\epsilon}\oint_S\big[k^2 G\,(\hat{\mathbf n}\times\boldsymbol H)+\big((\hat{\mathbf n}\times\boldsymbol H)\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf n}\times\boldsymbol E)\times\nabla' G\big]\,dS'$$

17.12 Vector Diffraction for Apertures

The Kirchhoff diffraction integral, Stratton-Chu, Kottler, and Franz formulas are equivalent only for a *closed* surface S.

If the surface is open, as in the case of an aperture, the four expressions in (17.10.13) and in (17.10.14) are no longer equivalent. In this case, the Kottler and Franz formulas remain equal to each other and give the correct expressions for the fields, in the sense that the resulting $\boldsymbol E(\mathbf r)$ and $\boldsymbol H(\mathbf r)$ satisfy Maxwell's equations [1113,1111,1125,1126].

For an open surface S bounded by a contour C, shown in Fig. 17.12.1, the Kottler and Franz formulas are related to the Stratton-Chu and the Kirchhoff diffraction integral formulas by the addition of some line-integral correction terms [1119]:

$$\boxed{\begin{aligned}\boldsymbol E(\mathbf r) &= \frac{1}{j\omega\epsilon}\int_S\big[k^2 G\,(\hat{\mathbf n}\times\boldsymbol H)+\big((\hat{\mathbf n}\times\boldsymbol H)\cdot\nabla'\big)\nabla' G + j\omega\epsilon\,(\hat{\mathbf n}\times\boldsymbol E)\times\nabla' G\big]\,dS'\\
&= \frac{1}{j\omega\epsilon}\,\nabla\times\Big(\nabla\times\int_S G\,(\hat{\mathbf n}\times\boldsymbol H)\,dS'\Big)+\nabla\times\int_S G\,(\hat{\mathbf n}\times\boldsymbol E)\,dS'\\
&= \int_S\big[-j\omega\mu\,G\,(\hat{\mathbf n}\times\boldsymbol H)+(\hat{\mathbf n}\cdot\boldsymbol E)\,\nabla' G + (\hat{\mathbf n}\times\boldsymbol E)\times\nabla' G\big]\,dS' - \frac{1}{j\omega\epsilon}\oint_C(\nabla' G)\,\boldsymbol H\cdot d\mathbf l\\
&= \int_S\Big[\boldsymbol E\,\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol E}{\partial n'}\Big]\,dS' - \oint_C G\,\boldsymbol E\times d\mathbf l - \frac{1}{j\omega\epsilon}\oint_C(\nabla' G)\,\boldsymbol H\cdot d\mathbf l\end{aligned}}\tag{17.12.1}$$

$$\boxed{\begin{aligned}\boldsymbol H(\mathbf r) &= \frac{1}{j\omega\mu}\int_S\big[-k^2 G\,(\hat{\mathbf n}\times\boldsymbol E)-\big((\hat{\mathbf n}\times\boldsymbol E)\cdot\nabla'\big)\nabla' G + j\omega\mu\,(\hat{\mathbf n}\times\boldsymbol H)\times\nabla' G\big]\,dS'\\
&= -\frac{1}{j\omega\mu}\,\nabla\times\Big(\nabla\times\int_S G\,(\hat{\mathbf n}\times\boldsymbol E)\,dS'\Big)+\nabla\times\int_S G\,(\hat{\mathbf n}\times\boldsymbol H)\,dS'\\
&= \int_S\big[j\omega\epsilon\,G\,(\hat{\mathbf n}\times\boldsymbol E)+(\hat{\mathbf n}\cdot\boldsymbol H)\,\nabla' G + (\hat{\mathbf n}\times\boldsymbol H)\times\nabla' G\big]\,dS' + \frac{1}{j\omega\mu}\oint_C(\nabla' G)\,\boldsymbol E\cdot d\mathbf l\\
&= -\int_S\Big[\boldsymbol H\,\frac{\partial G}{\partial n'} - G\,\frac{\partial \boldsymbol H}{\partial n'}\Big]\,dS' - \oint_C G\,\boldsymbol H\times d\mathbf l + \frac{1}{j\omega\mu}\oint_C(\nabla' G)\,\boldsymbol E\cdot d\mathbf l\end{aligned}}\tag{17.12.2}$$

The proof of the equivalence of these expressions is outlined in Problems 17.7 and 17.8. The Kottler-Franz formulas (17.12.1) and (17.12.2) are valid for points off the aperture surface S. The formulas are not consistent for points on the aperture. However, they have been used very successfully in practice to predict the radiation patterns of aperture antennas.

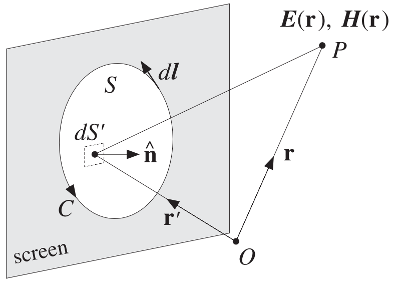


Fig. 17.12.1 Aperture surface S bounded by contour C.

The line-integral correction terms have a minor effect on the mainlobe and near sidelobes of the radiation pattern. Therefore, they can be ignored and the diffracted field can be calculated by any of the four alternative formulas, Kottler, Franz, Stratton-Chu, or Kirchhoff integral—all applied to the *open* surface S.

17.13 Fresnel Diffraction

In Sec. 17.4, we looked at the radiation fields arising from the Kottler-Franz formulas, where we applied the *Fraunhofer approximation* in which only linear phase variations over the aperture were kept in the propagation phase factor e^{-jkR}. Here, we consider the intermediate case of *Fresnel approximation* in which both linear and quadratic phase variations are retained.

We discuss the classical problem of diffraction of a spherical wave by a rectangular aperture, a slit, and a straight-edge using the Kirchhoff integral formula. The case of a plane wave incident on a conducting edge is discussed in Problem 17.11 using the field-equivalence principle and Kottler's formula and more accurately, in Sec. 17.15, using Sommerfeld's exact solution of the geometrical theory of diffraction. These examples are meant to be an introduction to the vast subject of diffraction.

In Fig. 17.13.1, we consider a rectangular aperture illuminated from the left by a point source radiating a spherical wave. We take the origin to be somewhere on the aperture plane, but eventually we will take it to be the point of intersection of the aperture plane and the line between the source and observation points P_1 and P_2.

The diffracted field at point P_2 may be calculated from the Kirchhoff formula applied to any of the cartesian components of the field:

$$E=\int_S\Big[E_1\,\frac{\partial G}{\partial n'} - G\,\frac{\partial E_1}{\partial n'}\Big]\,dS'\tag{17.13.1}$$

where E_1 is the spherical wave from the source point P_1 evaluated at the aperture point P', and G is the Green's function from P' to P_2:

$$E_1 = A_1\,\frac{e^{-jkR_1}}{R_1}\,,\quad G=\frac{e^{-jkR_2}}{4\pi R_2}\tag{17.13.2}$$

where A_1 is a constant. If $\mathbf r_1$ and $\mathbf r_2$ are the vectors pointing from the origin to the source and observation points, then we have for the distance vectors $\mathbf R_1$ and $\mathbf R_2$:

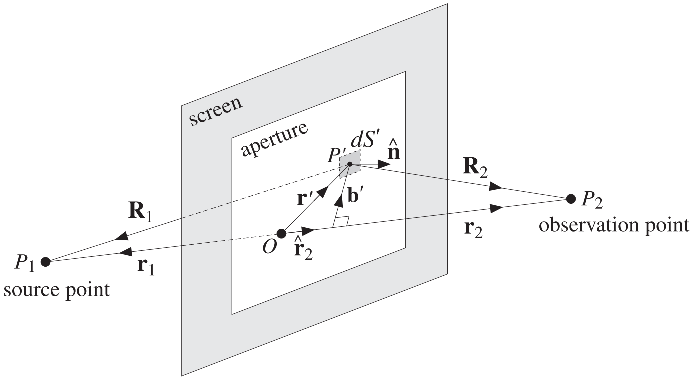


Fig. 17.13.1 Fresnel diffraction through rectangular aperture.

$$\begin{aligned}
\mathbf{R}_1 &= \mathbf{r}_1-\mathbf{r}', \quad R_1 = |\mathbf{r}_1-\mathbf{r}'| = \sqrt{r_1^2-2\mathbf{r}_1\cdot\mathbf{r}'+\mathbf{r}'\cdot\mathbf{r}'}\\
\mathbf{R}_2 &= \mathbf{r}_2-\mathbf{r}', \quad R_2 = |\mathbf{r}_2-\mathbf{r}'| = \sqrt{r_2^2-2\mathbf{r}_2\cdot\mathbf{r}'+\mathbf{r}'\cdot\mathbf{r}'}
\end{aligned} \qquad (17.13.3)$$

Therefore, the gradient operator ∇' can be written as follows when it acts on a function of $R_1 = |\mathbf{r}_1-\mathbf{r}'|$ or a function of $R_2 = |\mathbf{r}_2-\mathbf{r}'|$:

$$\nabla' = -\hat{\mathbf{R}}_1\frac{\partial}{\partial R_1}, \quad \nabla' = -\hat{\mathbf{R}}_2\frac{\partial}{\partial R_2}$$

where $\hat{\mathbf{R}}_1$ and $\hat{\mathbf{R}}_2$ are the unit vectors in the directions of $\mathbf{R}_1$ and $\mathbf{R}_2$. Thus, we have:

$$\begin{aligned}
\frac{\partial E_1}{\partial n'} &= \hat{\mathbf{n}}\cdot\nabla' E_1 = -\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_1\frac{\partial E_1}{\partial R_1} = (\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_1)\Big(jk+\frac{1}{R_1}\Big)A_1\frac{e^{-jkR_1}}{R_1}\\
\frac{\partial G}{\partial n'} &= \hat{\mathbf{n}}\cdot\nabla' G = -\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_2\frac{\partial G}{\partial R_2} = (\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_2)\Big(jk+\frac{1}{R_2}\Big)\frac{e^{-jkR_2}}{4\pi R_2}
\end{aligned} \qquad (17.13.4)$$

Dropping the $1/R^2$ terms, we find for the integrand of Eq. (17.13.1):

$$E_1\frac{\partial G}{\partial n'} - G\frac{\partial E_1}{\partial n'} = \frac{jkA_1}{4\pi R_1R_2}\big[(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{R}}_1)\big]e^{-jk(R_1+R_2)}$$

Except in the phase factor $e^{-jk(R_1+R_2)}$, we may replace $\mathbf{R}_1\simeq\mathbf{r}_1$ and $\mathbf{R}_2\simeq\mathbf{r}_2$, that is,

$$E_1\frac{\partial G}{\partial n'} - G\frac{\partial E_1}{\partial n'} = \frac{jkA_1}{4\pi r_1r_2}\big[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)\big]e^{-jk(R_1+R_2)} \qquad (17.13.5)$$

Thus, we have for the diffracted field at point P_2:

$$\boxed{E = \frac{jkA_1}{4\pi r_1r_2}\big[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)\big]\int_S e^{-jk(R_1+R_2)}\,dS'} \qquad (17.13.6)$$

The quantity $\big[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)\big]$ is an obliquity factor. Next, we set $r = r_1+r_2$ and define the "free-space" field at the point P_2:

$$E_0 = A_1\frac{e^{-jk(r_1+r_2)}}{r_1+r_2} = A_1\frac{e^{-jkr}}{r} \qquad (17.13.7)$$

If the origin were the point of intersection between the aperture plane and the line P_1P_2, then E_0 would represent the field received at point P_2 in the *unobstructed* case when the aperture and screen are absent.

The ratio $D = E/E_0$ may be called the *diffraction coefficient* and depends on the aperture and the relative geometry of the points P_1, P_2:

$$\boxed{D = \frac{E}{E_0} = \frac{jk}{4\pi F}\big[(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)-(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_1)\big]\int_S e^{-jk(R_1+R_2-r_1-r_2)}\,dS'} \qquad (17.13.8)$$

where we defined the "focal length" between r_1 and r_2:

$$\boxed{\frac{1}{F} = \frac{1}{r_1}+\frac{1}{r_2}} \quad\Rightarrow\quad \boxed{F = \frac{r_1r_2}{r_1+r_2}} \qquad (17.13.9)$$

The Fresnel approximation is obtained by expanding R_1 and R_2 in powers of r' and keeping only terms up to second order. We rewrite Eq. (17.13.3) in the form:

$$R_1 = r_1\sqrt{1-\frac{2\hat{\mathbf{r}}_1\cdot\mathbf{r}'}{r_1}+\frac{\mathbf{r}'\cdot\mathbf{r}'}{r_1^2}}, \quad R_2 = r_2\sqrt{1-\frac{2\hat{\mathbf{r}}_2\cdot\mathbf{r}'}{r_2}+\frac{\mathbf{r}'\cdot\mathbf{r}'}{r_2^2}}$$

Next, we apply the Taylor series expansion up to second order:

$$\sqrt{1+x} = 1+\frac{1}{2}x-\frac{1}{8}x^2$$

This gives the approximations of R_1, R_2, and $R_1+R_2-r_1-r_2$:

$$\begin{aligned}
R_1 &= r_1-\hat{\mathbf{r}}_1\cdot\mathbf{r}'+\frac{1}{2r_1}\big[\mathbf{r}'\cdot\mathbf{r}'-(\hat{\mathbf{r}}_1\cdot\mathbf{r}')^2\big]\\
R_2 &= r_2-\hat{\mathbf{r}}_2\cdot\mathbf{r}'+\frac{1}{2r_2}\big[\mathbf{r}'\cdot\mathbf{r}'-(\hat{\mathbf{r}}_2\cdot\mathbf{r}')^2\big]\\
R_1+R_2-r_1-r_2 &= -(\hat{\mathbf{r}}_1+\hat{\mathbf{r}}_2)\cdot\mathbf{r}'+\frac{1}{2}\bigg[\Big(\frac{1}{r_1}+\frac{1}{r_2}\Big)\mathbf{r}'\cdot\mathbf{r}'-\frac{(\hat{\mathbf{r}}_1\cdot\mathbf{r}')^2}{r_1}-\frac{(\hat{\mathbf{r}}_2\cdot\mathbf{r}')^2}{r_2}\bigg]
\end{aligned}$$

To simplify this expression, we now assume that the origin is the point of intersection of the line of sight P_1P_2 and the aperture plane. Then, the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are antiparallel and so are their unit vectors $\hat{\mathbf{r}}_1 = -\hat{\mathbf{r}}_2$. The linear terms cancel and the quadratic ones combine to give:

$$R_1+R_2-r_1-r_2 = \frac{1}{2F}\big[\mathbf{r}'\cdot\mathbf{r}'-(\hat{\mathbf{r}}_2\cdot\mathbf{r}')^2\big] = \frac{1}{2F}\,\big|\mathbf{r}'-\hat{\mathbf{r}}_2(\mathbf{r}'\cdot\hat{\mathbf{r}}_2)\big|^2 = \frac{1}{2F}\mathbf{b}'\cdot\mathbf{b}' \qquad (17.13.10)$$

where we defined $\mathbf{b}' = \mathbf{r}'-\hat{\mathbf{r}}_2(\mathbf{r}'\cdot\hat{\mathbf{r}}_2)$, which is the perpendicular vector from the point P' to the line-of-sight P_1P_2, as shown in Fig. 17.13.1.

It follows that the Fresnel approximation of the diffraction coefficient for an arbitrary aperture will be given by:

$$\boxed{D=\frac{E}{E_0}=\frac{jk(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2)}{2\pi F}\int_S e^{-jk(\mathbf{b}'\cdot\mathbf{b}')/(2F)}\,dS'} \tag{17.13.11}$$

A further simplification is obtained by assuming that the aperture plane is the xy-plane and that the line P_1P_2 lies on the yz plane at an angle θ with the z-axis, as shown in Fig. 17.13.2.

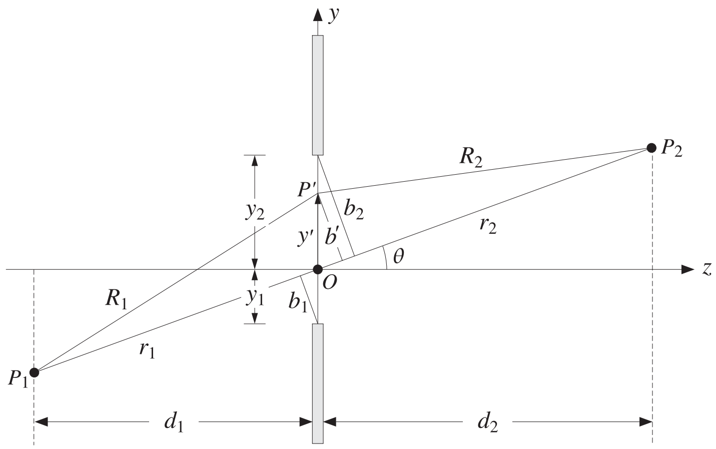


Fig. 17.13.2 Fresnel diffraction by rectangular aperture.

Then, we have $\mathbf{r}'=x'\hat{\mathbf{x}}+y'\hat{\mathbf{y}}$, $\hat{\mathbf{n}}=\hat{\mathbf{z}}$, and $\hat{\mathbf{r}}_2=\hat{\mathbf{z}}\cos\theta+\hat{\mathbf{y}}\sin\theta$. It follows that $\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_2=\cos\theta$, and the perpendicular distance $\mathbf{b}'\cdot\mathbf{b}'$ becomes:

$$\mathbf{b}'\cdot\mathbf{b}'=\mathbf{r}'\cdot\mathbf{r}'-(\hat{\mathbf{r}}'\cdot\hat{\mathbf{r}}_2)^2=x'^2+y'^2-(y'\sin\theta)^2=x'^2+y'^2\cos^2\theta$$

Then, the diffraction coefficient (17.13.11) becomes:

$$D=\frac{jk\cos\theta}{2\pi F}\int_{-x_1}^{x_2}\int_{-y1}^{y_2}e^{-jk(x'^2+y'^2\cos^2\theta)/2F}\,dx'dy' \tag{17.13.12}$$

where we assumed that the aperture limits are (with respect to the new origin):

$$-x_1\le x'\le x_2\,,\quad -y_1\le y'\le y_2$$

The end-points y_1,y_2 are shown in Fig. 17.13.2. The integrals may be expressed in terms of the Fresnel functions $C(x)$, $S(x)$, and $\mathcal{F}(x)=C(x)-jS(x)$ discussed in Appendix F. There, the complex function $\mathcal{F}(x)$ is defined by:

$$\mathcal{F}(x)=C(x)-jS(x)=\int_0^x e^{-j(\pi/2)u^2}\,du \tag{17.13.13}$$

We change integration variables to the normalized Fresnel variables:

$$u=\sqrt{\frac{k}{\pi F}}\,x'\,,\quad v=\sqrt{\frac{k}{\pi F}}\,y'\cos\theta \tag{17.13.14}$$

where $b'=y'\cos\theta$ is the perpendicular distance from P' to the line P_1P_2, as shown in Fig. 17.13.2. The corresponding end-points are:

$$u_i=\sqrt{\frac{k}{\pi F}}\,x_i\,,\quad v_i=\sqrt{\frac{k}{\pi F}}\,y_i\cos\theta=\sqrt{\frac{k}{\pi F}}\,b_i\,,\qquad i=1,2 \tag{17.13.15}$$

Note that the quantities $b_1=y_1\cos\theta$ and $b_2=y_2\cos\theta$ are the perpendicular distances from the edges to the line P_1P_2. Since $du\,dv=(k\cos\theta/\pi F)dx'dy'$, we obtain for the diffraction coefficient:

$$D=\frac{j}{2}\int_{-u_1}^{u_2}e^{-j\pi u^2/2}\,du\int_{-v_1}^{v_2}e^{-j\pi v^2/2}\,dv=\frac{j}{2}\big[\mathcal{F}(u_2)-\mathcal{F}(-u_1)\big]\big[\mathcal{F}(v_2)-\mathcal{F}(-v_1)\big]$$

Noting that $\mathcal{F}(x)$ is an odd function and that $j/2=1/(1-j)^2$, we obtain:

$$\boxed{D=\frac{E}{E_0}=\frac{\mathcal{F}(u_1)+\mathcal{F}(u_2)}{1-j}\,\frac{\mathcal{F}(v_1)+\mathcal{F}(v_2)}{1-j}}\quad\text{(rectangular aperture)} \tag{17.13.16}$$

The normalization factors $(1-j)$ correspond to the infinite aperture limit $u_1,u_2,v_1,v_2\to\infty$, that is, no aperture at all. Indeed, since the asymptotic value of $\mathcal{F}(x)$ is $\mathcal{F}(\infty)=(1-j)/2$, we have:

$$\frac{\mathcal{F}(u_1)+\mathcal{F}(u_2)}{1-j}\,\frac{\mathcal{F}(v_1)+\mathcal{F}(v_2)}{1-j}\longrightarrow\frac{\mathcal{F}(\infty)+\mathcal{F}(\infty)}{1-j}\,\frac{\mathcal{F}(\infty)+\mathcal{F}(\infty)}{1-j}=1$$

In the case of a *long slit* along the x-direction, we only take the limit $u_1,u_2\to\infty$:

$$\boxed{D=\frac{E}{E_0}=\frac{\mathcal{F}(v_1)+\mathcal{F}(v_2)}{1-j}}\quad\text{(diffraction by long slit)} \tag{17.13.17}$$

17.14 Knife-Edge Diffraction

The case of *straight-edge* or *knife-edge* diffraction is obtained by taking the limit $y_2\to\infty$, or $v_2\to\infty$, which corresponds to keeping the lower edge of the slit. In this limit $\mathcal{F}(v_2)\to\mathcal{F}(\infty)=(1-j)/2$. Denoting v_1 by v, we have:

$$\boxed{D(v)=\frac{1}{1-j}\left(\mathcal{F}(v)+\frac{1-j}{2}\right)}\,,\qquad v=\sqrt{\frac{k}{\pi F}}\,b_1 \tag{17.14.1}$$

Positive values of v correspond to positive values of the clearance distance b_1, placing the point P_2 in the illuminated region, as shown in Fig. 17.14.1. Negative values of v correspond to $b_1<0$, placing P_2 in the *geometrical shadow* region behind the edge.

The magnitude-square $|D|^2$ represents the intensity of the diffracted field relative to the intensity of the unobstructed field. Since $|1-j|^2=2$, we find:

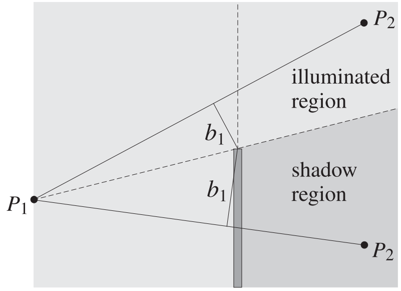


Fig. 17.14.1 Illuminated and shadow regions in straight-edge diffraction.

$$\boxed{|D(v)|^2 = \frac{|E|^2}{|E_0|^2} = \frac{1}{2}\left|\mathcal{F}(v)+\frac{1-j}{2}\right|^2} \qquad (17.14.2)$$

or, in terms of the real and imaginary parts of $\mathcal{F}(v)$:

$$\boxed{|D(v)|^2 = \frac{1}{2}\left[\left(C(v)+\frac{1}{2}\right)^2+\left(S(v)+\frac{1}{2}\right)^2\right]} \qquad (17.14.3)$$

The quantity $|D(v)|^2$ is plotted versus v in Fig. 17.14.2. At $v = 0$, corresponding to the line P_1P_2 grazing the top of the edge, we have $\mathcal{F}(0)=0$, $D(0)=1/2$, and $|D(0)|^2 = 1/4$ or a 6 dB loss. The first maximum in the illuminated region occurs at $v = 1.2172$ and has the value $|D(v)|^2 = 1.3704$, or a gain of 1.37 dB.

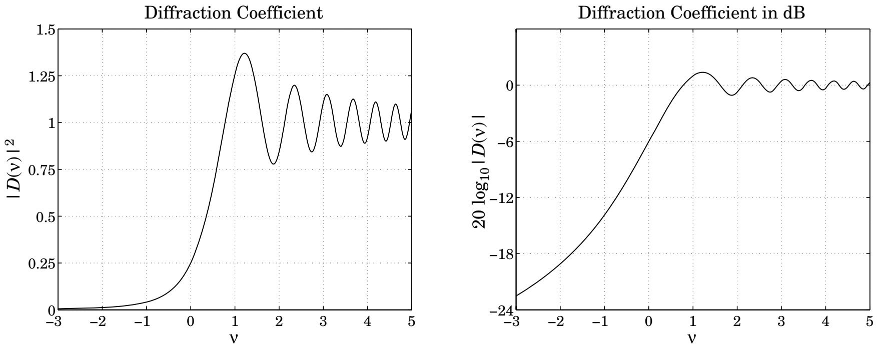


Fig. 17.14.2 Diffraction coefficient in absolute and dB units.

The asymptotic behavior of $D(v)$ for $v \to \pm\infty$ is obtained from Eq. (F.4). We have for large positive x:

$$\mathcal{F}(\pm x)\to \pm\left(\frac{1-j}{2}+\frac{j}{\pi x}\,e^{-j\pi x^2/2}\right)$$

This implies that:

$$\boxed{D(v) = \begin{cases} 1-\dfrac{1-j}{2\pi v}\,e^{-j\pi v^2/2}, & \text{for} \quad v\to+\infty \\[2ex] -\dfrac{1-j}{2\pi v}\,e^{-j\pi v^2/2}, & \text{for} \quad v\to-\infty \end{cases}} \qquad (17.14.4)$$

We may combine the two expressions into one with the help of the unit-step function $u(v)$ by writing $D(v)$ in the following form, which defines the asymptotic diffraction coefficient $d(v)$:

$$\boxed{D(v) = u(v)+d(v)\,e^{-j\pi v^2/2}} \qquad (17.14.5)$$

where $u(v)=1$ for $v\geq 0$ and $u(v)=0$ for $v<0$.

With $u(0)=1$, this definition requires $d(0)=D(0)-v(0)=0.5-1=-0.5$. But if we define $u(0)=0.5$, as is sometimes done, then, $d(0)=0$. The asymptotic behavior of $D(v)$ can now be expressed in terms of the asymptotic behavior of $d(v)$:

$$\boxed{d(v) = -\frac{1-j}{2\pi v}}, \qquad \text{for} \quad v\to\pm\infty \qquad (17.14.6)$$

In the illuminated region $D(v)$ tends to unity, whereas in the shadow region it decreases to zero with asymptotic dB attenuation or loss:

$$L = -10\log_{10}\left|d(v)\right|^2 = 10\log_{10}(2\pi^2v^2)\,, \quad \text{as} \quad v\to-\infty \qquad (17.14.7)$$

The MATLAB function `diffr` calculates the diffraction coefficient (17.14.1) at any vector of values of v. It has usage:

```
D = diffr(v);                    % knife-edge diffraction coefficient D(v)
```

For values $v\leq 0.7$, the diffraction loss can be approximated very well by the following function [1133]:

$$L = -10\log_{10}\left|D(v)\right|^2 = 6.9+20\log_{10}\left(\sqrt{(v+0.1)^2+1}-v-0.1\right) \qquad (17.14.8)$$

Example 17.14.1: *Diffraction Loss over Obstacles.* The propagation path loss over obstacles and irregular terrain is usually determined using knife-edge diffraction. Fig. 17.14.3 illustrates the case of two antennas communicating over an obstacle. For small angles θ, the focal length F is often approximated in several forms:

$$F = \frac{r_1r_2}{r_1+r_2}\simeq\frac{d_1d_2}{d_1+d_2}\simeq\frac{l_1l_2}{l_1+l_2}$$

These approximations are valid typically when d_1, d_2 are much greater than λ and the height h of the obstacle, typically, at least ten times greater. The clearance distance can be expressed in terms of the heights:

$$b_1 = y_1\cos\theta = \left(\frac{h_1d_2+h_2d_1}{d_1+d_2}-h\right)\cos\theta$$

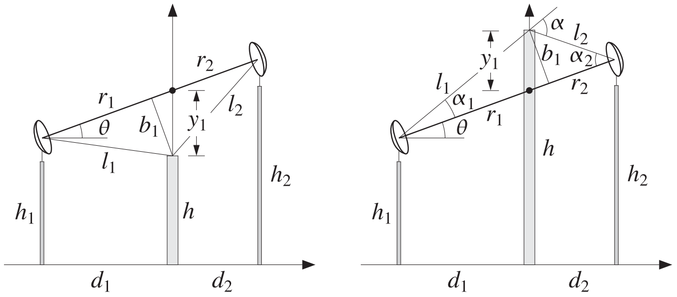


Fig. 17.14.3 Communicating antennas over an obstacle.

The distance b_1 can also be expressed approximately in terms of the subtended angles α_1, α_2, and α, shown in Fig. 17.14.3:

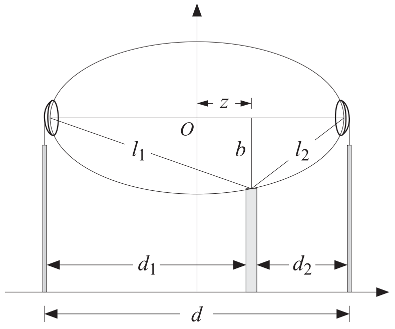
$$b_1 \simeq l_1\alpha_1 \simeq l_2\alpha_2 \quad \Rightarrow \quad b_1 = \sqrt{l_1 l_2 \alpha_1 \alpha_2} \tag{17.14.9}$$

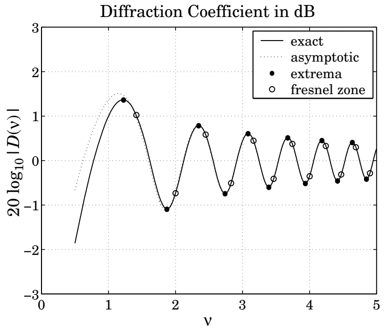
and in terms of α, we have:

$$\alpha_1 = \frac{\alpha l_2}{l_1 + l_2}\,, \quad \alpha_2 = \frac{\alpha l_1}{l_1 + l_2} \quad \Rightarrow \quad b_1 = \alpha F \quad \Rightarrow \quad \nu = \alpha\sqrt{\frac{2F}{\lambda}} \tag{17.14.10}$$

The case of multiple obstacles has been studied using appropriate modifications of the knife-edge diffraction problem and the geometrical theory of diffraction [1134–1147]. □

Example 17.14.2: *Fresnel Zones.* Consider two antennas separated by a distance d and an obstacle at distance z from the midpoint with clearance b, as shown below. Fresnel zones and the corresponding Fresnel zone ellipsoids help answer the question of what the minimum value of the clearance b should be for efficient communication between the antennas.





The diffraction coefficient $D(\nu)$ and its asymptotic form were given in Eqs. (17.14.1) and (17.14.4), that is,

$$D(\nu) = \frac{1}{1-j}\left(\mathcal{F}(\nu) + \frac{1-j}{2}\right), \quad \nu = \sqrt{\frac{k}{\pi F}}\,b = \sqrt{\frac{2}{\lambda F}}\,b\,, \quad F = \frac{d_1 d_2}{d_1 + d_2} \tag{17.14.11}$$

and for positive and large clearance b, or equivalently, for large positive ν,

$$D_{\text{as}}(\nu) = 1 - \frac{1-j}{2\pi\nu}\,e^{-j\pi\nu^2/2} = 1 - \frac{1}{\sqrt{2}\pi\nu}\,e^{-j\pi(\nu^2/2+1/4)} \tag{17.14.12}$$

As can be seen in the above figure on the right, the diffraction coefficients $D(\nu)$ and $D_{\text{as}}(\nu)$ agree closely even for small values of ν. Therefore, the extrema can be obtained from the asymptotic form. They correspond to the values of ν that cause the exponential in (17.14.12) to take on its extremal values of ± 1, that is, the ν's that satisfy $\nu^2/2 + 1/4 = n$, with integer n, or:

$$\nu_n = \sqrt{2n - 0.5}\,, \quad n = 1, 2, \dots \tag{17.14.13}$$

The corresponding values of $D(\nu)$, shown on the figure with black dots, are given by

$$D_{\text{as}}(\nu_n) = 1 - \frac{1}{\sqrt{2}\pi\nu_n}\,e^{-j\pi n} = 1 - \frac{1}{\sqrt{2}\pi\nu_n}\,(-1)^n \tag{17.14.14}$$

An alternative set of ν's, also corresponding to alternating almost extremum values, are those that define the conventional Fresnel zones, that is,

$$u_n = \sqrt{2n}\,, \quad n = 1, 2, \dots \tag{17.14.15}$$

These are indicated by open circles on the graph. The corresponding $D(\nu)$ values are:

$$D_{\text{as}}(u_n) = 1 - \frac{e^{-j\pi/4}}{\sqrt{2}\pi u_n}\,(-1)^n \tag{17.14.16}$$

For clearances b that correspond to ν's that are too small, i.e., $\nu < 0.5$, the diffraction coefficient $D(\nu)$ becomes too small, impeding efficient communication. The smallest acceptable clearance b is taken to correspond to the first maximum of $D(\nu)$, that is, $\nu = \nu_1$ or more simply $\nu = u_1 = \sqrt{2}$.

The locus of points (b, z) corresponding to a fixed value of ν, and hence to a fixed value of the diffraction coefficient $D(\nu)$, form an ellipsoid. This can be derived from (17.14.11) by setting $d_1 = d/2 + z$ and $d_2 = d/2 - z$, that is,

$$\nu = \sqrt{\frac{2}{\lambda F}}\,b \quad \Rightarrow \quad b^2 = \frac{\lambda F}{2}\,\nu^2 = \frac{\lambda(d^2/4 - z^2)}{2d}\,\nu^2\,, \quad \text{because } F = \frac{d_1 d_2}{d_1 + d_2} = \frac{d^2/4 - z^2}{d}$$

which can be rearranged into the equation of an ellipse:

$$\left(\frac{8}{\nu^2\lambda d}\right)b^2 + \left(\frac{4}{d^2}\right)z^2 = 1$$

For $\nu = u_1 = \sqrt{2}$, this defines the first Fresnel zone ellipse, which gives the minimum acceptable clearance for a given distance z:

$$\left(\frac{4}{\lambda d}\right)b^2 + \left(\frac{4}{d^2}\right)z^2 = 1 \tag{17.14.17}$$

If the obstacle is at midpoint ($z = 0$), the minimum clearance becomes:

$$b = \frac{1}{2}\sqrt{\lambda d} \tag{17.14.18}$$

For example, for a distance of $d = 1$ km, using a cell phone frequency of $f = 1$ GHz, corresponding to wavelength $\lambda = 30$ cm, we find $b = \sqrt{\lambda d}/2 = 8.66$ meters.

A common interpetation and derivation of Fresnel zones is to consider the path difference between the rays following the straight path connecting the two antennas and the path getting scattered from the obstacle, that is, $\Delta l = l_1 + l_2 - d$. From the indicated triangles, and assuming that $b \ll d_1$ and $b \ll d_2$, we find:

$$l_1 = \sqrt{d_1^2 + b^2} \simeq d_1 + \frac{b^2}{2d_1}\,, \quad l_2 = \sqrt{d_2^2 + b^2} \simeq d_2 + \frac{b^2}{2d_2}$$

which leads to the following path length Δl, expressed in terms of v:

$$\Delta l = l_1 + l_2 - d = \frac{b^2}{2}\left(\frac{1}{d_1} + \frac{1}{d_2}\right) = \frac{b^2}{2F} = \frac{\lambda}{4}\,v^2$$

The corresponding phase difference between the two paths, $e^{-jk\Delta l}$, will be then:

$$e^{-jk\Delta l} = e^{-j\pi v^2/2} \tag{17.14.19}$$

which has the same form as in the diffraction coefficient $D_{\text{as}}(v)$. The values $v = u_n = \sqrt{2n}$ will make the path difference a multiple of $\lambda/2$, that is, $\Delta l = n\lambda/2$, resulting in the alternating phase $e^{-jk\Delta l} = (-1)^n$.

The discrepancy between the choices v_n and u_n arises from using $D(v)$ to find the alternating maxima, versus using the plain phase (17.14.19). □

The Fresnel approximation is not invariant under shifting the origin. Our choice of origin above is not convenient because it depends on the observation point P_2. If we choose a fixed origin, such as the point O in Fig. 17.14.4, then, we must determine the corresponding Fresnel coefficient.

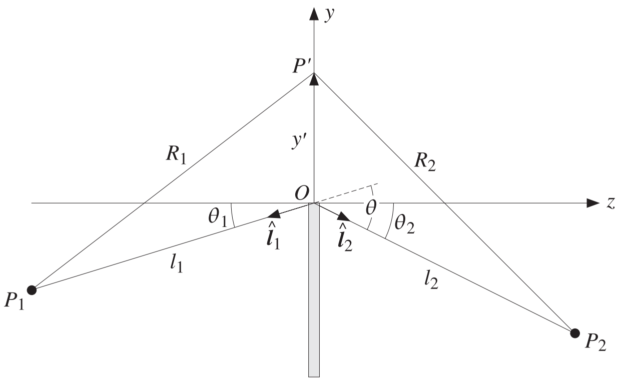


Fig. 17.14.4 Fresnel diffraction by straight edge.

We assume that the points P_1, P_2 lie on the yz plane and take P_2 to lie in the shadow region. The angles θ_1, θ_2 may be chosen to be positive or negative to obtain all possible locations of P_1, P_2 relative to the screen.

The diffraction coefficient is still given by Eq. (17.13.8) but with r_1, r_2 replaced by the distances l_1, l_2. The unit vectors towards P_1 and P_2 are:

$$\hat{\mathbf{l}}_1 = -\hat{\mathbf{z}}\cos\theta_1 - \hat{\mathbf{y}}\sin\theta_1\,, \qquad \hat{\mathbf{l}}_2 = \hat{\mathbf{z}}\cos\theta_2 - \hat{\mathbf{y}}\sin\theta_2 \tag{17.14.20}$$

Since $\mathbf{r}' = x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}}$ and $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, we find:

$$\hat{\mathbf{l}}_1\cdot\mathbf{r}' = -y'\sin\theta_1\,, \quad \hat{\mathbf{l}}_2\cdot\mathbf{r}' = -y'\sin\theta_2\,, \quad \hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_1 = -\cos\theta_1\,, \quad \hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_2 = \cos\theta_2$$

The quadratic approximation for the lengths R_1, R_2 gives, then:

$$\begin{aligned}
R_1 + R_2 - l_1 - l_2 &= -(\hat{\mathbf{l}}_1 + \hat{\mathbf{l}}_2)\cdot\mathbf{r}' + \frac{1}{2}\left[\left(\frac{1}{l_1} + \frac{1}{l_2}\right)(\mathbf{r}'\cdot\mathbf{r}') - \frac{(\hat{\mathbf{l}}_1\cdot\mathbf{r}')^2}{l_1} - \frac{(\hat{\mathbf{l}}_2\cdot\mathbf{r}')^2}{l_2}\right]\\
&= y'(\sin\theta_1 + \sin\theta_2) + \left(\frac{1}{l_1} + \frac{1}{l_2}\right)\frac{x'^2}{2} + \left(\frac{\cos^2\theta_1}{l_1} + \frac{\cos^2\theta_2}{l_2}\right)\frac{y'^2}{2}\\
&= \frac{1}{2F}x'^2 + \frac{1}{2F'}\left[y'^2 + 2F'y'(\sin\theta_1 + \sin\theta_2)\right]\\
&= \frac{1}{2F}x'^2 + \frac{1}{2F'}(y' + y_0)^2 - \frac{1}{2F'}\,y_0^2
\end{aligned}$$

where we defined the focal lengths F, F' and the shift y_0:

$$\frac{1}{F} = \frac{1}{l_1} + \frac{1}{l_2}\,, \quad \frac{1}{F'} = \frac{\cos^2\theta_1}{l_1} + \frac{\cos^2\theta_2}{l_2}\,, \quad y_0 = F'(\sin\theta_1 + \sin\theta_2) \tag{17.14.21}$$

Using these approximations in Eq. (17.13.6) and replacing r_1, r_2 by l_1, l_2, we find:

$$\begin{aligned}
E &= \frac{jkA_1 e^{-jk(l_1+l_2)}}{4\pi l_1 l_2}\left[(\hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_2) - (\hat{\mathbf{n}}\cdot\hat{\mathbf{l}}_1)\right]\int_S e^{-jk(R_1+R_2-l_1-l_2)}\,dS'\\
&= \frac{jkA_1 e^{-k(l_1+l_2)}}{4\pi l_1 l_2}(\cos\theta_1 + \cos\theta_2)\,e^{jky_0^2/2F'}\int e^{-jkx'^2/2F - jk(y'+y_0)^2/2F'}\,dx'dy'
\end{aligned}$$

The x'-integral is over the range $-\infty < x' < \infty$ and can be converted to a Fresnel integral with the change of variables $u = x'\sqrt{k/(\pi F)}$:

$$\int_{-\infty}^{\infty} e^{-jkx'^2/2F}\,dx' = \sqrt{\frac{\pi F}{k}}\int_{-\infty}^{\infty} e^{-j\pi u^2/2}\,du = \sqrt{\frac{\pi F}{k}}\,(1-j)$$

The y'-integral is over the upper-half of the xy-plane, that is, $0 \le y' < \infty$. Defining the Fresnel variables $u = (y' + y_0)\sqrt{k/(\pi F')}$ and $v = y_0\sqrt{k/(\pi F')}$, we find:

$$\int_0^{\infty} e^{-jk(y'+y_0)^2/2F'}\,dy' = \sqrt{\frac{\pi F'}{k}}\int_v^{\infty} e^{-j\pi u^2/2}\,du = \sqrt{\frac{\pi F'}{k}}\,(1-j)D(-v)$$

where the function $D(v)$ was defined in Eq. (17.14.1). Putting all the factors together, we may write the diffracted field at the point P_2 in the form:

$$\boxed{E = E_{\text{edge}}\frac{e^{-jkl_2}}{\sqrt{l_2}}\,D_{\text{edge}}} \qquad \text{(straight-edge diffraction)} \tag{17.14.22}$$

where we set $ky_0^2/2F' = \pi\nu^2/2$ and defined the incident field $E_{\rm edge}$ *at the edge* and the overall edge-diffraction coefficient $D_{\rm edge}$ by:

$$E_{\rm edge} = A_1\frac{e^{-jkl_1}}{l_1}\,,\qquad D_{\rm edge} = \sqrt{\frac{FF'}{l_2}}\left(\frac{\cos\theta_1+\cos\theta_2}{2}\right)e^{j\pi\nu^2/2}\,D(-\nu) \tag{17.14.23}$$

The second factor $(e^{-jkl_2}/\sqrt{l_2})$ in (17.14.22) may be interpreted as a *cylindrical wave* emanating from the edge as a result of the incident field $E_{\rm edge}$. The third factor $D_{\rm edge}$ is the angular gain of the cylindrical wave. The quantity ν may be written as:

$$\nu = \sqrt{\frac{k}{\pi F'}}\,y_0 = \sqrt{\frac{kF'}{\pi}}\,(\sin\theta_1+\sin\theta_2) \tag{17.14.24}$$

Depending on the sign and relative sizes of the angles θ_1 and θ_2, it follows that $\nu > 0$ when P_2 lies in the shadow region, and $\nu < 0$ when it lies in the illuminated region. For large positive ν, we may use Eq. (17.14.4) to obtain the asymptotic form of the edge-diffraction coefficient $D_{\rm edge}$:

$$D_{\rm edge} = \sqrt{\frac{FF'}{l_2}}\,\frac{\cos\theta_1+\cos\theta_2}{2}\,e^{j\pi\nu^2/2}\,\frac{1-j}{2\pi\nu}\,e^{-j\pi\nu^2/2} = \sqrt{\frac{FF'}{l_2}}\,\frac{\cos\theta_1+\cos\theta_2}{2}\,\frac{1-j}{2\pi\nu}$$

Writing $\sqrt{F/l_2} = \sqrt{l_1/(l_1+l_2)}$ and replacing ν from Eq. (17.14.24), the $\sqrt{F'}$ factor cancels and we obtain:

$$\boxed{D_{\rm edge} = \sqrt{\frac{l_1}{l_1+l_2}}\,\frac{(1-j)\,(\cos\theta_1+\cos\theta_2)}{4\sqrt{\pi k}\,(\sin\theta_1+\sin\theta_2)}} \tag{17.14.25}$$

This expression may be simplified further by defining the overall diffraction angle $\theta = \theta_1+\theta_2$, as shown in Fig. 17.14.4 and using the trigonometric identity:

$$\frac{\cos\theta_1+\cos\theta_2}{\sin\theta_1+\sin\theta_2} = \cot\left(\frac{\theta_1+\theta_2}{2}\right)$$

Then, Eq. (17.14.25) may be written in the form:

$$D_{\rm edge} = \sqrt{\frac{l_1}{l_1+l_2}}\,\frac{(1-j)}{4\sqrt{\pi k}}\,\cot\frac{\theta}{2} \tag{17.14.26}$$

The asymptotic diffraction coefficient is obtained from Eqs. (17.14.25) or (17.14.26) by taking the limit $l_1\to\infty$, which gives $\sqrt{l_1/(l_1+l_2)}\to 1$. Thus,

$$D_{\rm edge} = \frac{(1-j)\,(\cos\theta_1+\cos\theta_2)}{4\sqrt{\pi k}\,(\sin\theta_1+\sin\theta_2)} = \frac{(1-j)}{4\sqrt{\pi k}}\,\cot\frac{\theta}{2} \tag{17.14.27}$$

Eqs. (17.14.26) and (17.14.27) are equivalent to those given in [1125].

The two choices for the origin lead to two different expressions for the diffracted fields. However, the expressions agree near the forward direction, $\theta \simeq 0$. It is easily verified that both Eq. (17.14.1) and (17.14.26) lead to the same approximation for the diffracted field:

$$E = E_{\rm edge}\,\frac{e^{-jkl_2}}{\sqrt{l_2}}\sqrt{\frac{l_1}{l_1+l_2}}\,\frac{1-j}{2\sqrt{\pi k}\,\theta} \tag{17.14.28}$$

17.15 Geometrical Theory of Diffraction

Geometrical theory of diffraction is an extension of geometrical optics [1134–1140]. It views diffraction as a local edge effect. In addition to the ordinary rays of geometrical optics, it postulates the existence of "diffracted rays" from edges. The diffracted rays can reach into shadow regions, where geometrical optics fails.

An incident ray at an edge generates an infinity of diffracted rays emanating from the edge having different angular gains given by a diffraction coefficient $D_{\rm edge}$. An example of such a diffracted ray is given by Eq. (17.14.22).

The edge-diffraction coefficient $D_{\rm edge}$ depends on (a) the type of the incident wave, such as plane wave, or spherical, (b) the type and local geometry of the edge, such as a knife-edge or a wedge, and (c) the directions of the incident and diffracted rays.

The diffracted field and coefficient are usually taken to be in their asymptotic forms, like those of Eq. (17.15.26). The asymptotic forms are derived from certain exactly solvable canonical problems, such as a conducting edge, a wedge, and so on.

The first and most influential of all such problems was Sommerfeld's solution of a plane wave incident on a conducting half-plane [1111], and we discuss it below.

Fig. 17.15.1 shows a plane wave incident at an angle α on the conducting plane occupying half of the xz-plane for $x \ge 0$. The plane of incidence is taken to be the xy-plane. Because of the cylindrical symmetry of the problem, we may assume that there is no z-dependence and that the fields depend only on the cylindrical coordinates ρ,ϕ.

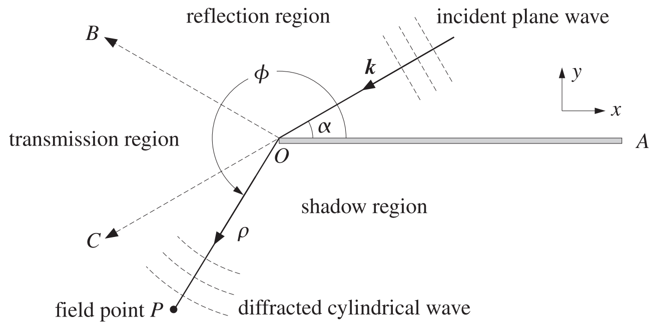


Fig. 17.15.1 Plane wave incident on conducting half-plane.

Two polarizations may be considered: TE, in which the electric field is $\boldsymbol{E} = \hat{\mathbf{z}}\,E_z$, and TM, which has $\boldsymbol{H} = \hat{\mathbf{z}}\,H_z$. Using cylindrical coordinates defined in Eq. (E.2) of Appendix E, and setting $\partial/\partial z = 0$, Maxwell's equations reduce in the two cases into:

$$\begin{aligned}
&\text{(TE)}\quad \nabla^2 E_z + k^2 E_z = 0, \quad H_\rho = -\frac{1}{j\omega\mu}\frac{1}{\rho}\frac{\partial E_z}{\partial\phi}\,, \quad H_\phi = \frac{1}{j\omega\mu}\frac{\partial E_z}{\partial\rho}\\
&\text{(TM)}\quad \nabla^2 H_z + k^2 H_z = 0, \quad E_\rho = \frac{1}{j\omega\epsilon}\frac{1}{\rho}\frac{\partial H_z}{\partial\phi}\,, \quad E_\phi = -\frac{1}{j\omega\epsilon}\frac{\partial H_z}{\partial\rho}
\end{aligned} \tag{17.15.1}$$

where $k^2 = \omega^2\mu\epsilon$, and the two-dimensional ∇^2 is in cylindrical coordinates:

$$\nabla^2 = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} \tag{17.15.2}$$

The boundary conditions require that the tangential electric field be zero on both sides of the conducting plane, that is, for $\phi = 0$ and $\phi = 2\pi$. In the TE case, the tangential electric field is E_z, and in the TM case, $E_x = E_\rho\cos\phi - E_\phi\sin\phi = E_\rho = (1/j\omega\epsilon\rho)(\partial H_z/\partial\phi)$, for $\phi = 0, 2\pi$. Thus, the boundary conditions are:

$$\begin{aligned} &(\text{TE}) \quad E_z = 0, &&\text{for } \phi = 0 \text{ and } \phi = 2\pi \\ &(\text{TM}) \quad \frac{\partial H_z}{\partial\phi} = 0, &&\text{for } \phi = 0 \text{ and } \phi = 2\pi \end{aligned} \tag{17.15.3}$$

In Fig. 17.15.1, we assume that $0 \le \alpha \le 90^{\mathrm{o}}$ and distinguish three wedge regions defined by the half-plane and the directions along the reflected and transmitted rays:

$$\begin{aligned} &\text{reflection region } (AOB): && 0 \le \phi \le \pi - \alpha \\ &\text{transmission region } (BOC): && \pi - \alpha \le \phi \le \pi + \alpha \\ &\text{shadow region } (COA): && \pi + \alpha \le \phi \le 2\pi \end{aligned} \tag{17.15.4}$$

The case when $90^{\mathrm{o}} \le \alpha \le 180^{\mathrm{o}}$ is shown in Fig. 17.15.2, in which α has been redefined to still be in the range $0 \le \alpha \le 90^{\mathrm{o}}$. The three wedge regions are now:

$$\begin{aligned} &\text{reflection region } (AOB): && 0 \le \phi \le \alpha \\ &\text{transmission region } (BOC): && \alpha \le \phi \le 2\pi - \alpha \\ &\text{shadow region } (COA): && 2\pi - \alpha \le \phi \le 2\pi \end{aligned} \tag{17.15.5}$$

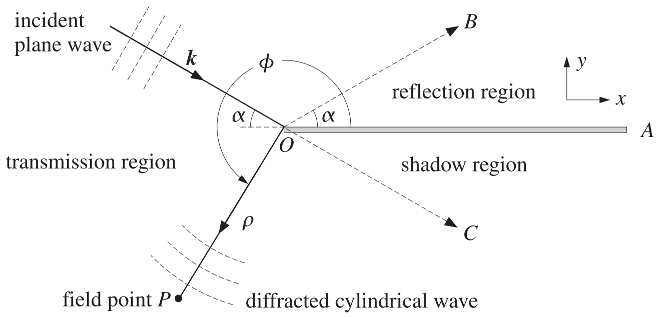


Fig. 17.15.2 Plane wave incident on conducting half-plane.

We construct the Sommerfeld solution in stages. We start by looking for solutions of the Helmholtz equation $\nabla^2 U + k^2 U = 0$ that have the factored form: $U = ED$, where E is also a solution, but a simple one, such as that of the incident plane wave. Using the differential identities of Appendix C, we have:

$$\nabla^2 U + k^2 U = D(\nabla^2 E + k^2 E) + E\nabla^2 D + 2\,\nabla E\cdot\nabla D$$

Thus, the conditions $\nabla^2 U + k^2 U = 0$ and $\nabla^2 E + k^2 E = 0$ require:

$$E\nabla^2 D + 2\,\nabla E\cdot\nabla D = 0 \quad\Rightarrow\quad \nabla^2 D + 2(\nabla\ln E)\cdot\nabla D = 0 \tag{17.15.6}$$

If we assume that E is of the form $E = e^{jf}$, where f is a real-valued function, then, equating to zero the real and imaginary parts of $\nabla^2 E + k^2 E = 0$, we find for f:

$$\nabla^2 E + k^2 E = E(k^2 - \nabla f\cdot\nabla f + j\nabla^2 f) = 0 \quad\Rightarrow\quad \nabla^2 f = 0\,, \quad \nabla f\cdot\nabla f = k^2 \tag{17.15.7}$$

Next, we assume that D is of the form:

$$D = D_0\int_{-\infty}^{v} e^{-jg(u)}du \tag{17.15.8}$$

where D_0 is a constant, v is a function of ρ, ϕ, and $g(u)$ is a real-valued function to be determined. Noting that $\nabla D = D_0 e^{-jg}\,\nabla v$ and $\nabla g = g'(v)\,\nabla v$, we find:

$$\nabla D = D_0 e^{-jg}\,\nabla v\,, \quad \nabla^2 D = D_0 e^{-jg}\big(\nabla^2 v - jg'(v)\,\nabla v\cdot\nabla v\big)$$

Then, it follows from Eq. (17.15.6) that $\nabla^2 D + 2(\nabla\ln E)\cdot\nabla D = \nabla^2 D + j\nabla f\cdot\nabla D$ and:

$$\nabla^2 D + j\nabla f\cdot\nabla D = D_0 e^{-jg}\big[\nabla^2 v + j(2\,\nabla f\cdot\nabla v - g'\nabla v\cdot\nabla v)\big] = 0$$

Equating the real and imaginary parts to zero, we obtain the two conditions:

$$\nabla^2 v = 0\,, \quad \frac{2\,\nabla f\cdot\nabla v}{\nabla v\cdot\nabla v} = g'(v) \tag{17.15.9}$$

Sommerfeld's solution involves the Fresnel diffraction coefficient of Eq. (17.14.1), which can be written as follows:

$$\boxed{D(v) = \frac{1}{1-j}\left[\frac{1-j}{2} + \mathcal{F}(v)\right] = \frac{1}{1-j}\int_{-\infty}^{v} e^{-j\pi u^2/2}du} \tag{17.15.10}$$

Therefore, we are led to choose $g(u) = \pi u^2/2$ and $D_0 = 1/(1-j)$. To summarize, we may construct a solution of the Helmholtz equation in the form:

$$\nabla^2 U + k^2 U = 0\,, \qquad \boxed{U = ED = e^{jf}D(v)} \tag{17.15.11}$$

where f and v must be chosen to satisfy the four conditions:

$$\begin{aligned} &\nabla^2 f = 0, && \nabla f\cdot\nabla f = k^2 \\ &\nabla^2 v = 0, && \frac{2\,\nabla f\cdot\nabla v}{\nabla v\cdot\nabla v} = g'(v) = \pi v \end{aligned} \tag{17.15.12}$$

It can be verified easily that the functions $u = \rho^a\cos a\phi$ and $u = \rho^a\sin a\phi$ are solutions of the two-dimensional Laplace equation $\nabla^2 u = 0$, for *any* value of the parameter a. Taking f to be of the form $f = A\rho^a\cos a\phi$, we have the condition:

$$\nabla f = Aa\rho^{a-1}\big[\hat{\boldsymbol{\rho}}\cos a\phi - \hat{\boldsymbol{\phi}}\sin a\phi\big] \quad\Rightarrow\quad \nabla f\cdot\nabla f = A^2a^2\rho^{2(a-1)} = k^2$$

This immediately implies that $a = 1$ and $A^2 = k^2$, so that $A = \pm k$. Thus, $f = A\rho\cos\phi = \pm k\rho\cos\phi$. Next, we choose $\nu = B\rho^a\cos a\phi$. Then:

$$
\begin{aligned}
\nabla f &= A(\hat{\boldsymbol{\rho}}\cos\phi - \hat{\boldsymbol{\phi}}\sin\phi)\\
\nabla\nu &= Ba\rho^{a-1}\big[\hat{\boldsymbol{\rho}}\cos a\phi - \hat{\boldsymbol{\phi}}\sin a\phi\big]\\
\nabla f\cdot\nabla\nu &= ABa\rho^{a-1}\big[\cos\phi\cos a\phi + \sin\phi\sin a\phi\big] = ABa\rho^{a-1}\cos(\phi - a\phi)\\
\nabla\nu\cdot\nabla\nu &= B^2a^2\rho^{2(a-1)}
\end{aligned}
$$

Then, the last of the conditions (17.15.12) requires that:

$$\frac{1}{\pi\nu}\,\frac{2\,\nabla f\cdot\nabla\nu}{\nabla\nu\cdot\nabla\nu} = \frac{2A\rho^{1-2a}\cos(\phi - a\phi)}{\pi a B^2\cos a\phi} = 1$$

which implies that $a = 1/2$ and $B^2 = 2A/\pi a = 4A/\pi$. But since $A = \pm k$, only the case $A = k$ is compatible with a real coefficient B. Thus, we have $B^2 = 4k/\pi$, or, $B = \pm 2\sqrt{k/\pi}$.

In a similar fashion, we find that if we take $\nu = B\rho^a\sin a\phi$, then $a = 1/2$, but now $B^2 = -4A/\pi$, requiring that $A = -k$, and $B = \pm 2\sqrt{k/\pi}$. In summary, we have the following solutions of the conditions (17.15.12):

$$
\begin{aligned}
f &= +k\rho\cos\phi\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2}\\
f &= -k\rho\cos\phi\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2}
\end{aligned}
\tag{17.15.13}
$$

The corresponding solutions (17.15.11) of the Helmholtz equation are:

$$
\boxed{
\begin{aligned}
U(\rho,\phi) &= e^{jk\rho\cos\phi}\,D(\nu)\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2}\\
U(\rho,\phi) &= e^{-jk\rho\cos\phi}\,D(\nu)\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2}
\end{aligned}
}
\tag{17.15.14}
$$

The function $D(\nu)$ may be replaced by the equivalent form of Eq. (17.14.5) in order to bring out its asymptotic behavior for large ν:

$$
\begin{aligned}
U(\rho,\phi) &= e^{jk\rho\cos\phi}\big[u(\nu) + d(\nu)e^{-j\pi\nu^2/2}\big], &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2}\\
U(\rho,\phi) &= e^{-jk\rho\cos\phi}\big[u(\nu) + d(\nu)e^{-j\pi\nu^2/2}\big], &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2}
\end{aligned}
$$

Using the trigonometric identities $\cos\phi = 2\cos^2(\phi/2) - 1 = 1 - 2\sin^2(\phi/2)$, we find for the two choices of ν:

$$
\begin{aligned}
k\rho\cos\phi - \frac{1}{2}\pi\nu^2 &= k\rho\left[\cos\phi - 2\cos^2\frac{\phi}{2}\right] = -k\rho\\
-k\rho\cos\phi - \frac{1}{2}\pi\nu^2 &= -k\rho\left[\cos\phi + 2\sin^2\frac{\phi}{2}\right] = -k\rho
\end{aligned}
$$

Thus, an alternative form of Eq. (17.15.14) is:

$$
\boxed{
\begin{aligned}
U(\rho,\phi) &= e^{jk\rho\cos\phi}\,u(\nu) + e^{-jk\rho}\,d(\nu)\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi}{2}\\
U(\rho,\phi) &= e^{-jk\rho\cos\phi}\,u(\nu) + e^{-jk\rho}\,d(\nu)\,, &\quad \nu &= \pm 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi}{2}
\end{aligned}
}
\tag{17.15.15}
$$

Shifting the origin of the angle ϕ still leads to a solution. Indeed, defining $\phi' = \phi \pm \alpha$, we note the property $\partial/\partial\phi' = \partial/\partial\phi$, which implies the invariance of the Laplace operator under this change. The functions $U(\rho, \phi\pm\alpha)$ are the elementary solutions from which the Sommerfeld solution is built.

Considering the TE case first, the incident plane wave in Fig. 17.15.1 is $\boldsymbol{E} = \hat{\mathbf{z}}\,E_i$, where $E_i = E_0e^{-j\boldsymbol{k}\cdot\mathbf{r}}$, with $\mathbf{r} = \hat{\mathbf{x}}\,\rho\cos\phi + \hat{\mathbf{y}}\,\rho\sin\phi$ and $\boldsymbol{k} = -k(\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$. It follows that:

$$
\begin{aligned}
\boldsymbol{k}\cdot\mathbf{r} &= -k\rho(\cos\phi\cos\alpha + \sin\phi\sin\alpha) = -k\rho\cos(\phi - \alpha)\\
E_i &= E_0e^{-j\boldsymbol{k}\cdot\mathbf{r}} = E_0e^{jk\rho\cos(\phi-\alpha)}
\end{aligned}
\tag{17.15.16}
$$

The image of this electric field with respect to the perfect conducting plane will be the reflected field $E_r = -E_0e^{-j\boldsymbol{k}_r\cdot\mathbf{r}}$, where $\boldsymbol{k}_r = k(-\hat{\mathbf{x}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$, resulting in $E_r = -E_0e^{jk\rho\cos(\phi+\alpha)}$. The sum $E_i + E_r$ does vanish for $\phi = 0$ and $\phi = 2\pi$, but it also vanishes for $\phi = \pi$. Therefore, it is an appropriate solution for a full conducting plane (the entire xz-plane), not for the half-plane.

Sommerfeld's solution, which satisfies the correct boundary conditions, is obtained by forming the linear combinations of the solutions of the type of Eq. (17.15.14):

$$\boxed{E_z = E_0\big[e^{jk\rho\cos\phi_i}\,D(\nu_i) - e^{jk\rho\cos\phi_r}\,D(\nu_r)\big]}\quad\text{(TE)}\tag{17.15.17}$$

where

$$
\begin{aligned}
\phi_i &= \phi - \alpha\,, &\quad \nu_i &= 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi_i}{2}\\
\phi_r &= \phi + \alpha\,, &\quad \nu_r &= 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\cos\frac{\phi_r}{2}
\end{aligned}
\tag{17.15.18}
$$

For the TM case, we form the sum instead of the difference:

$$\boxed{H_z = H_0\big[e^{jk\rho\cos\phi_i}\,D(\nu_i) + e^{jk\rho\cos\phi_r}\,D(\nu_r)\big]}\quad\text{(TM)}\tag{17.15.19}$$

The boundary conditions (17.15.3) are satisfied by both the TE and TM solutions. As we see below, the choice of the positive sign in the definitions of ν_i and ν_r was required in order to produce the proper diffracted field in the shadow region. Using the alternative forms (17.15.15), we separate the terms of the solution as follows:

$$\boxed{E_z = E_0e^{jk\rho\cos\phi_i}\,u(\nu_i) - E_0e^{jk\rho\cos\phi_r}\,u(\nu_r) + E_0e^{-jk\rho}\big[d(\nu_i) - d(\nu_r)\big]}\tag{17.15.20}$$

The first two terms correspond to the incident and reflected fields. The third term is the diffracted field. The algebraic signs of ν_i and ν_r are as follows within the reflection, transmission, and shadow regions of Eq. (17.15.4):

$$\begin{array}{lll} \text{reflection region:} & 0 \le \phi < \pi - \alpha, & \nu_i > 0, \quad \nu_r > 0 \\ \text{transmission region:} & \pi - \alpha < \phi < \pi + \alpha, & \nu_i > 0, \quad \nu_r < 0 \\ \text{shadow region:} & \pi + \alpha < \phi \le 2\pi, & \nu_i < 0, \quad \nu_r < 0 \end{array} \qquad (17.15.21)$$

The unit-step functions will be accordingly present or absent resulting in the following fields in these three regions:

$$\boxed{\begin{array}{ll} \text{reflection region:} & E_z = E_i + E_r + E_d \\ \text{transmission region:} & E_z = E_i + E_d \\ \text{shadow region:} & E_z = E_d \end{array}} \qquad (17.15.22)$$

where we defined the incident, reflected, and diffracted fields:

$$\begin{aligned} E_i &= E_0 e^{jk\rho\cos\phi_i} \\ E_r &= -E_0 e^{jk\rho\cos\phi_r} \\ E_d &= E_0 e^{-jk\rho}\big[d(\nu_i) - d(\nu_r)\big] \end{aligned} \qquad (17.15.23)$$

The diffracted field is present in all three regions, and in particular it is the only one in the shadow region. For large ν_i and ν_r (positive or negative), we may replace $d(\nu)$ by its asymptotic form $d(\nu) = -(1-j)/(2\pi\nu)$ of Eq. (17.14.6), resulting in the asymptotic diffracted field:

$$\begin{aligned} E_d &= -E_0 e^{-jk\rho}\,\frac{1-j}{2\pi}\left(\frac{1}{\nu_i} - \frac{1}{\nu_r}\right) \\ &= -E_0 e^{-jk\rho}\,\frac{1-j}{2\pi 2\sqrt{k/\pi}\,\rho^{1/2}}\left(\frac{1}{\cos(\phi_i/2)} - \frac{1}{\cos(\phi_r/2)}\right) \end{aligned}$$

which can be written in the form:

$$\boxed{E_d = E_0\,\frac{e^{-jk\rho}}{\rho^{1/2}}\,D_{\text{edge}}} \qquad (17.15.24)$$

with an edge-diffraction coefficient:

$$\boxed{D_{\text{edge}} = -\frac{1-j}{4\sqrt{\pi k}}\left(\frac{1}{\cos\dfrac{\phi_i}{2}} - \frac{1}{\cos\dfrac{\phi_r}{2}}\right)} \qquad (17.15.25)$$

Using a trigonometric identity, we may write D_{edge} as follows:

$$D_{\text{edge}} = -\frac{1-j}{4\sqrt{\pi k}}\left(\frac{1}{\cos\dfrac{\phi-\alpha}{2}} - \frac{1}{\cos\dfrac{\phi+\alpha}{2}}\right) = -\frac{1-j}{\sqrt{\pi k}}\,\frac{\sin\dfrac{\phi}{2}\sin\dfrac{\alpha}{2}}{\cos\phi + \cos\alpha} \qquad (17.15.26)$$

Eqs. (17.15.22) and (17.15.24) capture the essence of the geometrical theory of diffraction: In addition to the ordinary incident and reflected geometric optics rays, one also has diffracted rays in all directions corresponding to a cylindrical wave emanating from the edge with a directional gain of D_{edge}.

For the case of Fig. 17.15.2, the incident and reflected plane waves have propagation vectors $\boldsymbol{k} = k(\hat{\mathbf{z}}\cos\alpha - \hat{\mathbf{y}}\sin\alpha)$ and $\boldsymbol{k}_r = k(\hat{\mathbf{z}}\cos\alpha + \hat{\mathbf{y}}\sin\alpha)$. These correspond to the incident and reflected fields:

$$E_i = E_0 e^{-j\boldsymbol{k}\cdot\mathbf{r}} = E_0 e^{-jk\rho\cos(\phi+\alpha)}\,, \quad E_r = -E_0 e^{-j\boldsymbol{k}_r\cdot\mathbf{r}} = -E_0 e^{-jk\rho\cos(\phi-\alpha)}$$

In this case, the Sommerfeld TE and TM solutions take the form:

$$\boxed{\begin{aligned} E_z &= E_0\big[e^{-jk\rho\cos\phi_i}D(\nu_i) - e^{-jk\rho\cos\phi_r}D(\nu_r)\big] \\ H_z &= H_0\big[e^{-jk\rho\cos\phi_i}D(\nu_i) + e^{-jk\rho\cos\phi_r}D(\nu_r)\big] \end{aligned}} \qquad (17.15.27)$$

where, now:

$$\begin{aligned} \phi_i &= \phi + \alpha\,, & \nu_i &= 2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi_i}{2} \\ \phi_r &= \phi - \alpha\,, & \nu_r &= -2\sqrt{\frac{k}{\pi}}\,\rho^{1/2}\sin\frac{\phi_r}{2} \end{aligned} \qquad (17.15.28)$$

The choice of signs in ν_i and ν_r are such that they are both negative within the shadow region defined by Eq. (17.15.5). The same solution can also be obtained from Fig. 17.15.1 and Eq. (17.15.17) by replacing α by $\pi - \alpha$.

17.16 Rayleigh-Sommerfeld Diffraction Theory

In this section, we recast Kirchhoff's diffraction formula in a form that uses a Dirichlet Green's function (i.e., one that vanishes on the boundary surface) and obtain the Rayleigh-Sommerfeld diffraction formula. In the next section, we show that this reformulation is equivalent to the plane-wave spectrum approach to diffraction, and in Sec. 17.18, we use it to obtain the usual Fresnel and Fraunhofer approximations and discuss a few applications from Fourier optics.

We will work with the scalar case, but the same method can be used for the vector case. With reference to Fig. 17.16.1, we we consider a scalar field $E(\mathbf{r})$ that satisfies the source-free Helmholtz equation, $(\nabla^2 + k^2)E(\mathbf{r}) = 0$, over the right half-space $z \ge 0$.

We consider a closed surface consisting of the surface S_∞ of a sphere of very large radius centered at the observation point $\mathbf{r}$ and bounded on the left by its intersection S with the xy plane, as shown in the Fig. 17.16.1. Clearly, in the limit of infinite radius, the volume V bounded by $S + S_\infty$ is the right half-space $z \ge 0$, and S becomes the entire xy plane. Applying Eq. (17.10.3) to volume V, we have:

$$\int_V \big[G(\nabla'^2 E + k^2 E) - E(\nabla'^2 G + k^2 G)\big]\,dV' = -\oint_{S+S_\infty}\left[G\,\frac{\partial E}{\partial n'} - E\,\frac{\partial G}{\partial n'}\right]dS' \quad (17.16.1)$$

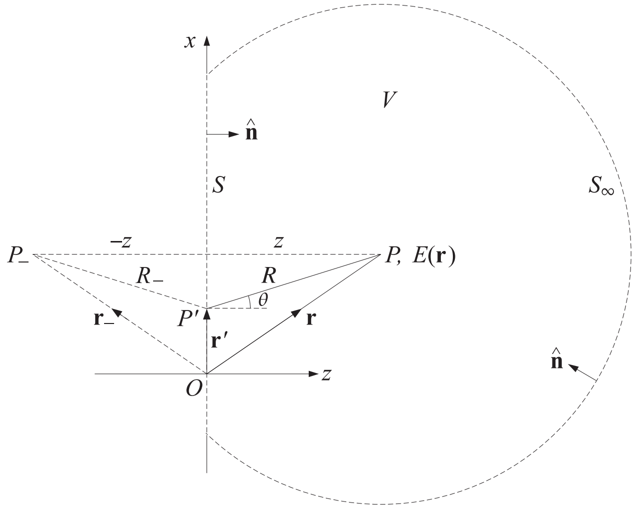


Fig. 17.16.1 Fields determined from their values on the *xy*-plane surface.

The surface integral over S_∞ can be ignored by noting that $\hat{\mathbf{n}}$ is the negative of the radial unit vector and therefore, we have after adding and subtracting the term $jkEG$:

$$-\oint_{S_\infty}\left[G\,\frac{\partial E}{\partial n'}-E\,\frac{\partial G}{\partial n'}\right]dS'=\oint_{S_\infty}\left[G\left(\frac{\partial E}{\partial r}+jkE\right)-E\left(\frac{\partial G}{\partial r}+jkG\right)\right]dS'$$

Assuming Sommerfeld's outgoing radiation condition:

$$r\left(\frac{\partial E}{\partial r}+jkE\right)\to 0\,,\quad \text{as}\quad r\to\infty$$

and noting that $G=e^{-jkr}/4\pi r$ also satisfies the same condition, it follows that the above surface integral vanishes in the limit of large radius r. Then, in the notation of Eq. (17.10.4), we obtain the standard Kirchhoff diffraction formula:

$$E(\mathbf{r})\,u_V(\mathbf{r})=\oint_S\left[E\,\frac{\partial G}{\partial n'}-G\,\frac{\partial E}{\partial n'}\right]dS' \tag{17.16.2}$$

Thus, if $\mathbf{r}$ lies in the right half-space, the left-hand side will be equal to $E(\mathbf{r})$, and if $\mathbf{r}$ is in the left half-space, it will vanish. Given a point $\mathbf{r}=(x,y,z)$, we define its reflection relative to the xy plane by $\mathbf{r}_-=(x,y,-z)$. The distance between $\mathbf{r}_-$ and a source point $\mathbf{r}'=(x',y',z')$ can be written in terms of the distance between the original point $\mathbf{r}$ and the reflected source point $\mathbf{r}'_-=(x',y',-z')$:

$$R_-=|\mathbf{r}_--\mathbf{r}'|=\sqrt{(x-x')^2+(y-y')^2+(z+z')^2}=|\mathbf{r}-\mathbf{r}'_-|$$

whereas

$$R=|\mathbf{r}-\mathbf{r}'|=\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}$$

This leads us to define the reflected Green's function:

$$G_-(\mathbf{r},\mathbf{r}')=\frac{e^{-jkR_-}}{4\pi R_-}=G(\mathbf{r}-\mathbf{r}'_-)=G(\mathbf{r}_--\mathbf{r}') \tag{17.16.3}$$

and the Dirichlet Green's function:

$$G_d(\mathbf{r},\mathbf{r}')=G(\mathbf{r},\mathbf{r}')-G_-(\mathbf{r},\mathbf{r}')=\frac{e^{-jkR}}{4\pi R}-\frac{e^{-jkR_-}}{4\pi R_-} \tag{17.16.4}$$

For convenience, we may choose the origin to lie on the xy plane. Then, as shown in Fig. 17.16.1, when the source point $\mathbf{r}'$ lies on the xy plane (i.e., $z'=0$), the function $G_d(\mathbf{r},\mathbf{r}')$ will vanish because $R=R_-$. Next, we apply Eq. (17.16.1) at the observation point $\mathbf{r}$ in the right half-space and at its reflection in the left half-plane, where (17.16.1) vanishes:

$$\begin{aligned}E(\mathbf{r})&=\oint_S\left[E\,\frac{\partial G}{\partial n'}-G\,\frac{\partial E}{\partial n'}\right]dS'\,,&&\text{at point }\mathbf{r}\\0&=\oint_S\left[E\,\frac{\partial G_-}{\partial n'}-G_-\,\frac{\partial E}{\partial n'}\right]dS'\,,&&\text{at point }\mathbf{r}_-\end{aligned}$$

where G_- stands for $G(\mathbf{r}_--\mathbf{r}')$. But on the xy plane boundary, $G_-=G$ so that if we subtract the two expressions we may eliminate the term $\partial E/\partial n'$, which is the reason for using the Dirichlet Green's function:[†]

$$E(\mathbf{r})=\oint_S E(\mathbf{r}')\,\frac{\partial}{\partial n'}(G-G_-)\,dS'=\oint_S E(\mathbf{r}')\,\frac{\partial G_d}{\partial n'}\,dS'$$

On the xy plane, we have $\hat{\mathbf{n}}=\hat{\mathbf{z}}$, and therefore

$$\frac{\partial G}{\partial n'}=\left.\frac{\partial G}{\partial z'}\right|_{z'=0}\quad\text{and}\quad\frac{\partial G_-}{\partial n'}=\left.\frac{\partial G_-}{\partial z'}\right|_{z'=0}=-\left.\frac{\partial G}{\partial z'}\right|_{z'=0}$$

Then, the two derivative terms double resulting in the *Rayleigh-Sommerfeld* diffraction formula [1110,1111]:

$$\boxed{E(\mathbf{r})=2\oint_S E(\mathbf{r}')\,\frac{\partial G}{\partial z'}\,dS'}\quad\text{(Rayleigh-Sommerfeld)} \tag{17.16.5}$$

The indicated derivative of G can be expressed as follows:

$$\left.\frac{\partial G}{\partial z'}\right|_{z'=0}=\frac{z}{R}\left(jk+\frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R}=\cos\theta\left(jk+\frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R} \tag{17.16.6}$$

where θ is the angle between the z-axis and the direction between the source and observation points, as shown in Fig. 17.16.1. For distances $R\gg\lambda$, or equivalently, $k\gg 1/R$, one obtains the approximation:

$$\left.\frac{\partial G}{\partial z'}\right|_{z'=0}=jk\cos\theta\,\frac{e^{-jkR}}{4\pi R}\,,\quad\text{for}\quad R\gg\lambda \tag{17.16.7}$$

[†] By adding instead of subtracting the above integrals, we obtain the alternative Green's function $G_s=G+G_-$, having vanishing derivative on the boundary.

This approximation will be used in Sec. 17.18 to obtain the standard Fresnel diffraction representation. The quantity $\cos\theta = z/R$ is an "obliquity" factor and is usually omitted for *paraxial* observation points that are near the z axis.

Equation (17.16.5) expresses the field at any point in the right half-space in terms of its values on the xy plane. In the practical application of this result, if the plane consists of an infinite opaque screen with an aperture S cut in it, then the integration in (17.16.5) is restricted only over the aperture S. The usual *Kirchhoff approximations* assume that: (a) the field is zero over the opaque screen, and (b) the field, $E(\mathbf{r}')$, over the aperture is equal to the incident field from the left.

Eq. (17.16.5) is also valid in the vectorial case for each component of the electric field $\boldsymbol{E}(\mathbf{r})$. However, these components are not independent of each other since they must satisfy $\nabla\cdot\boldsymbol{E} = 0$, and are also coupled to the magnetic field through Maxwell's equations. Taking into account these constraints, one arrives at a modified form of (17.16.5). We pursue this further in the next section.

17.17 Plane-Wave Spectrum Representation

The plane-wave spectrum representation builds up a (single-frequency) propagating wave $E(x,y,z)$ as a linear combination of plane waves $e^{-j(k_x x+k_y y+k_z z)}$. The only assumption is that the field must satisfy the wave equation, which for harmonic time dependence $e^{j\omega t}$ is the Helmholtz equation

$$(\nabla^2 + k^2)E(x,y,z) = 0\,, \quad k = \frac{\omega}{c} \tag{17.17.1}$$

where c is the speed of light in the propagation medium (assumed here to be homogeneous, isotropic, and lossless.) In solving the Helmholtz equation, one assumes initially a solution of the form:

$$E(x,y,z) = \hat{E}(k_x,k_y,z)\,e^{-jk_x x}e^{-jk_y y}$$

Inserting this into Eq. (17.17.1) and replacing $\partial_x \to -jk_x$ and $\partial_y \to -jk_y$, we obtain:

$$\left(-k_x^2 - k_y^2 + \frac{\partial^2}{\partial z^2} + k^2\right)\hat{E}(k_x,k_y,z) = 0$$

or, defining $k_z^2 = k^2 - k_x^2 - k_y^2$, we have:

$$\frac{\partial^2\hat{E}(k_x,k_y,z)}{\partial z^2} = -(k^2 - k_x^2 - k_y^2)\hat{E}(k_x,k_y,z) = -k_z^2\,\hat{E}(k_x,k_y,z)$$

Its solution describing *forward-moving* waves ($z \geq 0$) is:

$$\boxed{\hat{E}(k_x,k_y,z) = \hat{E}(k_x,k_y,0)\,e^{-jk_z z}} \tag{17.17.2}$$

If $k_x^2 + k_y^2 < k^2$, the wavenumber k_z is real-valued and the solution describes a *propagating* wave. If $k_x^2 + k_y^2 > k^2$, then k_z is imaginary and the solution describes an *evanescent* wave decaying with distance z. The two cases can be combined into one by defining k_z as follows:

$$\boxed{k_z = \begin{cases} \sqrt{k^2 - k_x^2 - k_y^2}\,, & \text{if} \quad k_x^2 + k_y^2 \leq k^2 \\ -j\sqrt{k_x^2 + k_y^2 - k^2}\,, & \text{if} \quad k_x^2 + k_y^2 > k^2 \end{cases}} \tag{17.17.3}$$

In the latter case, we have the decaying solution:

$$\hat{E}(k_x,k_y,z) = \hat{E}(k_x,k_y,0)\,e^{-z\sqrt{k_x^2+k_y^2-k^2}}\,, \quad z \geq 0$$

The complete space dependence is $\hat{E}(k_x,k_y,0)\,e^{-jk_x x - jk_y y}e^{-jk_z z}$. The most general solution of Eq. (17.17.1) is obtained by adding up such plane-waves, that is, by the *spatial* two-dimensional inverse Fourier transform:

$$\boxed{E(x,y,z) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} \hat{E}(k_x,k_y,0)\,e^{-jk_x x - jk_y y}e^{-jk_z z}\,\frac{dk_x\,dk_y}{(2\pi)^2}} \tag{17.17.4}$$

This is the *plane-wave spectrum representation.* Because k_z is given by Eq. (17.17.3), this solution is composed, in general, by both propagating and evanescent modes. Of course, for large z, only the propagating modes survive. Setting $z = 0$, we recognize $\hat{E}(k_x,k_y,0)$ to be the spatial Fourier transform of the field, $E(x,y,0)$, on the xy plane:

$$\begin{aligned} E(x,y,0) &= \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} \hat{E}(k_x,k_y,0)\,e^{-jk_x x - jk_y y}\,\frac{dk_x\,dk_y}{(2\pi)^2} \\ \hat{E}(k_x,k_y,0) &= \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} E(x,y,0)\,e^{jk_x x + jk_y y}\,dx\,dy \end{aligned} \tag{17.17.5}$$

As in Chap. 3, we may give a system-theoretic interpretation to these results. Defining the "propagation" spatial filter $\hat{g}(k_x,k_y,z) = e^{-jk_z z}$, then Eq. (17.17.2) reads:

$$\hat{E}(k_x,k_y,z) = \hat{g}(k_x,k_y,z)\hat{E}(k_x,k_y,0) \tag{17.17.6}$$

This multiplicative relationship in the wavenumber domain translates into a convolutional equation in the space domain. Denoting by $g(x,y,z)$ the spatial inverse Fourier transform of $\hat{g}(k_x,k_y,z) = e^{-jk_z z}$, that is,

$$\boxed{g(x,y,z) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} e^{-jk_x x - jk_y y}e^{-jk_z z}\,\frac{dk_x\,dk_y}{(2\pi)^2}} \tag{17.17.7}$$

we may write Eq. (17.17.4) in the form:

$$\boxed{E(x,y,z) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} E(x',y',0)\,g(x-x',y-y',z)\,dx'\,dy'} \tag{17.17.8}$$

Eq. (17.17.8) is *equivalent* to the Rayleigh-Sommerfeld formula (17.16.5). Indeed, it follows from Eq. (D.19) of Appendix D that

$$g(x-x',y-y',z) = -2\frac{\partial G}{\partial z} = 2\frac{\partial G}{\partial z'}\,, \quad G = \frac{e^{-jkR}}{4\pi R}\,, \quad R = |\mathbf{r}-\mathbf{r}'| \tag{17.17.9}$$

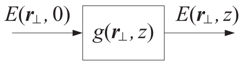
with the understanding that $z' = 0$. Thus, (17.17.8) takes the form of (17.16.5).

Next, we discuss the vector case as it applies to electromagnetic fields. To simplify the notation, we define the two-dimensional transverse vectors $\boldsymbol{r}_\perp = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$ and $\boldsymbol{k}_\perp = \hat{\mathbf{x}}k_x + \hat{\mathbf{y}}k_y$, as well as the transverse gradient $\nabla_\perp = \hat{\mathbf{x}}\,\partial_x + \hat{\mathbf{y}}\,\partial_y$, so that the full three-dimensional gradient is

$$\nabla = \hat{\mathbf{x}}\,\partial_x + +\hat{\mathbf{y}}\,\partial_y + \hat{\mathbf{z}}\,\partial_z = \nabla_\perp + \hat{\mathbf{z}}\,\partial_z$$

In this notation, Eq. (17.17.6) takes the form $\hat{E}(\boldsymbol{k}_\perp, z) = \hat{g}(\boldsymbol{k}_\perp, z)\hat{E}(\boldsymbol{k}_\perp, 0)$. The plane-wave spectrum representations (17.17.4) and (17.17.8) now read (where the integral sign denotes double integration):

$$\boxed{\begin{aligned} E(\boldsymbol{r}_\perp, z) &= \int_{-\infty}^{\infty} \hat{E}(\boldsymbol{k}_\perp, 0)\, e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \\ &= \int_{-\infty}^{\infty} E(\boldsymbol{r}_\perp', 0)\, g(\boldsymbol{r}_\perp - \boldsymbol{r}_\perp', z)\, d^2\boldsymbol{r}_\perp' \end{aligned}} \tag{17.17.10}$$


$E(\boldsymbol{r}_\perp, 0)$
$g(\boldsymbol{r}_\perp, z)$
$E(\boldsymbol{r}_\perp, z)$

and

$$g(\boldsymbol{r}_\perp, z) = \int_{-\infty}^{\infty} e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \tag{17.17.11}$$

In the vectorial case, $E(\boldsymbol{r}_\perp, z)$ is replaced by a three-dimensional field, which can be decomposed into its transverse x, y components and its longitudinal part along z:

$$\boldsymbol{E} = \hat{\mathbf{x}}E_x + \hat{\mathbf{y}}E_y + \hat{\mathbf{z}}E_z \equiv \boldsymbol{E}_\perp + \hat{\mathbf{z}}E_z$$

The Rayleigh-Sommerfeld and plane-wave spectrum representations apply separately to each component and can be written vectorially as

$$\boldsymbol{E}(\boldsymbol{r}_\perp, z) = \int_{-\infty}^{\infty} \hat{\boldsymbol{E}}(\boldsymbol{k}_\perp, 0)\, e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} = \int_{-\infty}^{\infty} \boldsymbol{E}(\boldsymbol{r}_\perp', 0)\, g(\boldsymbol{r}_\perp - \boldsymbol{r}_\perp', z)\, d^2\boldsymbol{r}_\perp' \tag{17.17.12}$$

Because $\boldsymbol{E}$ must satisfy the source-free Gauss's law, $\nabla\cdot\boldsymbol{E} = 0$, this imposes certain constraints among the Fourier components $\hat{\boldsymbol{E}}$ that must be taken into account in writing (17.17.12). Indeed, we have from (17.17.12)

$$\nabla\cdot\boldsymbol{E} = -j\int_{-\infty}^{\infty} \boldsymbol{k}\cdot\hat{\boldsymbol{E}}(\boldsymbol{k}_\perp, 0)\, e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} = 0$$

which requires that $\boldsymbol{k}\cdot\hat{\boldsymbol{E}}(\boldsymbol{k}_\perp, 0) = 0$. Separating the transverse and longitudinal parts, we have:

$$\boldsymbol{k}\cdot\hat{\boldsymbol{E}} = \boldsymbol{k}_\perp\cdot\hat{\boldsymbol{E}}_\perp + k_z\hat{E}_z = 0 \quad\Rightarrow\quad \hat{E}_z = -\frac{\boldsymbol{k}_\perp\cdot\hat{\boldsymbol{E}}_\perp}{k_z}$$

It follows that the Fourier vector $\hat{\boldsymbol{E}}$ must have the form:

$$\hat{\boldsymbol{E}} = \hat{\boldsymbol{E}}_\perp + \hat{\mathbf{z}}\hat{E}_z = \hat{\boldsymbol{E}}_\perp - \hat{\mathbf{z}}\,\frac{\boldsymbol{k}_\perp\cdot\hat{\boldsymbol{E}}_\perp}{k_z} \tag{17.17.13}$$

and, therefore, it is expressible only in terms of its transverse components $\hat{\boldsymbol{E}}_\perp$. Then, the correct plane-wave spectrum representation (17.17.12) becomes:

$$\boldsymbol{E}(\boldsymbol{r}_\perp, z) = \int_{-\infty}^{\infty}\left(\hat{\boldsymbol{E}}_\perp(\boldsymbol{k}_\perp, 0) - \hat{\mathbf{z}}\,\frac{\boldsymbol{k}_\perp\cdot\hat{\boldsymbol{E}}_\perp(\boldsymbol{k}_\perp, 0)}{k_z}\right) e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \tag{17.17.14}$$

But from the Weyl representations (D.18) and (D.20), we have with $G = e^{-jkr}/4\pi r$:

$$-2\frac{\partial G}{\partial z} = \int_{-\infty}^{\infty} e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2}\,, \qquad -2\nabla_\perp G = \int_{-\infty}^{\infty} \frac{\boldsymbol{k}_\perp}{k_z}\, e^{-jk_z z}\, e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2}$$

Then, (17.17.14) can be written convolutionally in the form:

$$\boldsymbol{E}(\boldsymbol{r}_\perp, z) = -2\int\left[\boldsymbol{E}_\perp\,\frac{\partial G}{\partial z} - \hat{\mathbf{z}}\,\nabla_\perp G\cdot\boldsymbol{E}_\perp\right] d^2\boldsymbol{r}_\perp' \tag{17.17.15}$$

where here $G = e^{-jkR}/4\pi R$ with $R = |\mathbf{r} - \mathbf{r}'|$ and $z' = 0$, and $\boldsymbol{E}_\perp$ in the integrand stands for $\boldsymbol{E}_\perp(\boldsymbol{r}_\perp', 0)$. Eq. (17.17.15) follows from the observation that in (17.17.14) the following products of Fourier transforms (in $\boldsymbol{k}_\perp$) appear, which become convolutions in the $\boldsymbol{r}_\perp$ domain:

$$\hat{\boldsymbol{E}}_\perp(\boldsymbol{k}_\perp, 0)\cdot e^{-jk_z z} \quad\text{and}\quad \hat{\boldsymbol{E}}_\perp(\boldsymbol{k}_\perp, 0)\cdot\left(\frac{\boldsymbol{k}_\perp}{k_z}\, e^{-jk_z z}\right)$$

Because $\boldsymbol{E}_\perp(\boldsymbol{r}_\perp', 0)$ does not depend on $\mathbf{r}$, it is straightforward to verify using some vector identities that

$$\hat{\mathbf{z}}\,\nabla_\perp G\cdot\boldsymbol{E}_\perp - \boldsymbol{E}_\perp\,\frac{\partial G}{\partial z} = \nabla\times(\hat{\mathbf{z}}\times\boldsymbol{E}_\perp G)$$

This gives rise to the Rayleigh-Sommerfeld-type equation for the vector case:

$$\boldsymbol{E}(\boldsymbol{r}_\perp, z) = 2\nabla\times\int \hat{\mathbf{z}}\times\boldsymbol{E}_\perp(\boldsymbol{r}_\perp', 0)\, G(R)\, d^2\boldsymbol{r}_\perp' \tag{17.17.16}$$

which can be abbreviated as

$$\boxed{\boldsymbol{E}(\mathbf{r}) = 2\,\nabla\times\int_S \hat{\mathbf{z}}\times\boldsymbol{E}_\perp G\, dS'} \tag{17.17.17}$$

The magnetic field can be determined from Faraday's law, $\nabla\times\boldsymbol{E} = -j\omega\mu\boldsymbol{H}$:

$$\boxed{\boldsymbol{H}(\mathbf{r}) = \frac{2}{-j\omega\mu}\,\nabla\times\boldsymbol{E} = \frac{2}{-j\omega\mu}\,\nabla\times\left(\nabla\times\int_S \hat{\mathbf{z}}\times\boldsymbol{E}_\perp G\, dS'\right)} \tag{17.17.18}$$

The same results can be derived more directly by using the Franz formulas (17.10.13) and making use of the extinction theorem as we did in Sec. 17.16. Applying (17.10.13) to the closed surface $S + S_\infty$ of Fig. 17.16.1, and dropping the S_∞ term, it follows that the left-hand side of (17.10.13) will be zero if the point $\mathbf{r}$ is not in the right half-space. To simplify the notation, we define the vectors:

$$\boldsymbol{e} = \int_S G\,(\hat{\mathbf{z}}\times\boldsymbol{E})\, dS'\,, \quad \boldsymbol{h} = \int_S G\,(\hat{\mathbf{z}}\times\boldsymbol{H})\, dS'$$

where we took S to be the xy plane with the unit vector $\hat{\mathbf{n}} = \hat{\mathbf{z}}$. Then, Eqs. (17.10.13) and (17.10.14) can be written as:

$$\boldsymbol{E}(\mathbf{r}) = \frac{1}{j\omega\epsilon}\nabla\times(\nabla\times\boldsymbol{h}) + \nabla\times\boldsymbol{e}\,, \quad \boldsymbol{H}(\mathbf{r}) = \frac{1}{-j\omega\mu}\nabla\times(\nabla\times\boldsymbol{e}) + \nabla\times\boldsymbol{h}$$

Noting that $\boldsymbol{e}, \boldsymbol{h}$ are transverse vectors and using some vector identities and the decomposition $\nabla = \nabla_\perp + \hat{\mathbf{z}}\,\partial_z$, we can rewrite the above in a form that explicitly separates the transverse and longitudinal parts, so that if $\mathbf{r}$ is in the right half-space:

$$\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= \frac{1}{j\omega\epsilon}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{h}) - \partial_z^2\boldsymbol{h} + \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{h})\big] + \nabla_\perp\times\boldsymbol{e} + \hat{\mathbf{z}}\times\partial_z\boldsymbol{e} \\
\boldsymbol{H}(\mathbf{r}) &= \frac{1}{-j\omega\mu}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{e}) - \partial_z^2\boldsymbol{e} + \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{e})\big] + \nabla_\perp\times\boldsymbol{h} + \hat{\mathbf{z}}\times\partial_z\boldsymbol{h}
\end{aligned} \tag{17.17.19}$$

If $\mathbf{r}$ is chosen to be the reflected point $\mathbf{r}_-$ on the left half-space, then $G_- = G$ and the vectors $\boldsymbol{e}, \boldsymbol{h}$ remain the same, but the gradient with respect to $\mathbf{r}_-$ is now $\nabla_- = \nabla_\perp - \hat{\mathbf{z}}\,\partial_z$, arising from the replacement $z \to -z$. Thus, replacing $\partial_z \to -\partial_z$ in (17.17.19) and setting the result to zero, we have:

$$\begin{aligned}
0 &= \frac{1}{j\omega\epsilon}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{h}) - \partial_z^2\boldsymbol{h} - \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{h})\big] + \nabla_\perp\times\boldsymbol{e} - \hat{\mathbf{z}}\times\partial_z\boldsymbol{e} \\
0 &= \frac{1}{-j\omega\mu}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{e}) - \partial_z^2\boldsymbol{e} - \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{e})\big] + \nabla_\perp\times\boldsymbol{h} - \hat{\mathbf{z}}\times\partial_z\boldsymbol{h}
\end{aligned} \tag{17.17.20}$$

Separating (17.17.20) into its transverse and longitudinal parts, we have:

$$\begin{aligned}
&\frac{1}{j\omega\epsilon}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{h}) - \partial_z^2\boldsymbol{h}\big] = \hat{\mathbf{z}}\times\partial_z\boldsymbol{e}\,, \quad \frac{1}{j\omega\epsilon}\big[\hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{h})\big] = \nabla_\perp\times\boldsymbol{e} \\
&\frac{1}{-j\omega\mu}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{e}) - \partial_z^2\boldsymbol{e}\big] = \hat{\mathbf{z}}\times\partial_z\boldsymbol{h}\,, \quad \frac{1}{-j\omega\mu}\big[\hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{e}\big] = \nabla_\perp\times\boldsymbol{h}
\end{aligned} \tag{17.17.21}$$

Using these conditions into Eq. (17.17.19), we obtain the doubling of terms:

$$\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= 2\,\nabla_\perp\times\boldsymbol{e} + 2\,\hat{\mathbf{z}}\times\partial_z\boldsymbol{e} = 2\,\nabla\times\boldsymbol{e} \\
\boldsymbol{H}(\mathbf{r}) &= \frac{2}{-j\omega\mu}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{e}) - \partial_z^2\boldsymbol{e} + \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{e})\big] = \frac{2}{-j\omega\mu}\nabla\times(\nabla\times\boldsymbol{e})
\end{aligned} \tag{17.17.22}$$

which are the same as Eqs. (17.17.17) and (17.17.18). Alternatively, we may express the diffracted fields in terms of the values of the magnetic field at the xy surface:

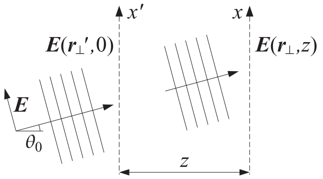
$$\begin{aligned}
\boldsymbol{E}(\mathbf{r}) &= \frac{2}{j\omega\epsilon}\big[\nabla_\perp\times(\nabla_\perp\times\boldsymbol{h}) - \partial_z^2\boldsymbol{h} + \hat{\mathbf{z}}\,\partial_z(\nabla_\perp\cdot\boldsymbol{h})\big] = \frac{2}{j\omega\epsilon}\nabla\times(\nabla\times\boldsymbol{h}) \\
\boldsymbol{H}(\mathbf{r}) &= 2\,\nabla_\perp\times\boldsymbol{h} + 2\,\hat{\mathbf{z}}\times\partial_z\boldsymbol{h} = 2\,\nabla\times\boldsymbol{h}
\end{aligned} \tag{17.17.23}$$

Eqs. (17.17.17) and (17.17.23) are equivalent to applying the Franz formulas with the field-equivalent surface currents of Eqs. (17.1.2) and (17.1.3), respectively.

As in the scalar case, the vector method is applied in practice by assuming that the fields on the aperture plane are the same as the incident fields from the left.

Example 17.17.1: *Oblique Plane Wave.* Here, we show that the plane-wave spectrum method correctly generates an ordinary plane wave from its transverse values at an input plane. Consider a TM electromagnetic wave propagating at an angle θ_0 with respect to the z axis, as shown in the figure below. The electric field at an arbitrary point, and its transverse part evaluated on the plane $z' = 0$, are given by

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}_\perp, z) &= E_0\,(\hat{\mathbf{x}}\cos\theta_0 - \hat{\mathbf{z}}\sin\theta_0)\,e^{-j(k_x^0 x + k_z^0 z)} \\
\boldsymbol{E}_\perp(\boldsymbol{r}_\perp', 0) &= \hat{\mathbf{x}}\,E_0\cos\theta_0\,e^{-jk_x^0 x'} = \hat{\mathbf{x}}\,E_0\cos\theta_0\,e^{-j\boldsymbol{k}_\perp^0\cdot\boldsymbol{r}_\perp'} \\
k_x^0 &= k\sin\theta_0\,, \quad k_y^0 = 0\,, \quad k_z^0 = k\cos\theta_0 \\
\boldsymbol{k}_\perp^0 &= \hat{\mathbf{x}}\,k_x^0 + \hat{\mathbf{y}}\,k_y^0 = \hat{\mathbf{x}}\,k\sin\theta_0
\end{aligned}$$



It follows that the spatial Fourier transform of $\boldsymbol{E}_\perp(\boldsymbol{r}_\perp', 0)$ will be

$$\hat{\boldsymbol{E}}_\perp(\boldsymbol{k}_\perp, 0) = \int_{-\infty}^{\infty}\hat{\mathbf{x}}\,E_0\cos\theta_0\,e^{-j\boldsymbol{k}_\perp^0\cdot\boldsymbol{r}_\perp'}\,e^{j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp'}\,d^2\boldsymbol{r}_\perp' = \hat{\mathbf{x}}\,E_0\cos\theta_0\,(2\pi)^2\delta(\boldsymbol{k}_\perp - \boldsymbol{k}_\perp^0)$$

Then, the integrand of Eq. (17.17.14) becomes

$$\hat{\boldsymbol{E}}_\perp - \hat{\mathbf{z}}\,\frac{\boldsymbol{k}_\perp\cdot\hat{\boldsymbol{E}}_\perp}{k_z} = E_0\,(\hat{\mathbf{x}}\cos\theta_0 - \hat{\mathbf{z}}\sin\theta_0)\,(2\pi)^2\delta(\boldsymbol{k}_\perp - \boldsymbol{k}_\perp^0)$$

and Eq. (17.17.14) gives

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{r}_\perp, z) &= \int_{-\infty}^{\infty} E_0\,(\hat{\mathbf{x}}\cos\theta_0 - \hat{\mathbf{z}}\sin\theta_0)\,(2\pi)^2\delta(\boldsymbol{k}_\perp - \boldsymbol{k}_\perp^0)\,e^{-jk_z z}\,e^{-j\boldsymbol{k}_\perp\cdot\boldsymbol{r}_\perp}\,\frac{d^2\boldsymbol{k}_\perp}{(2\pi)^2} \\
&= E_0\,(\hat{\mathbf{x}}\cos\theta_0 - \hat{\mathbf{z}}\sin\theta_0)\,e^{-j(k_x^0 x + k_z^0 z)}
\end{aligned}$$

which is the correct expression for the plane wave. For a TE wave a similar result holds. □

17.18 Fresnel Diffraction and Fourier Optics

The Fresnel approximation for planar apertures is obtained from the Rayleigh-Sommerfeld formula (17.16.5). Using (17.16.6), we have:

$$E(\boldsymbol{r}_\perp, z) = \int_S E(\boldsymbol{r}_\perp', 0)\,\frac{2z}{R}\left(jk + \frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R}\,d^2\boldsymbol{r}_\perp' \tag{17.18.1}$$

where $R = \sqrt{(x-x')^2 + (y-y')^2 + z^2} = \sqrt{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2 + z^2}$. The Fresnel approximation assumes that z is large enough such that $|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'| \ll z$, which can realized if the aperture has dimension d so that $|\boldsymbol{r}_\perp'| < d$, and one assumes that the observation point $\boldsymbol{r}_\perp$ remains close to the z-axis (the paraxial approximation) such that $|\boldsymbol{r}_\perp| < d$, and z is chosen such that $z \gg d$. Then, we can approximate R as follows:

$$R = \sqrt{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2 + z^2} = z\sqrt{1 + \frac{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2}{z^2}} \simeq z\left[1 + \frac{1}{2}\,\frac{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2}{z^2}\right] = z + \frac{|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2}{2z}$$

where we used the Taylor series expansion $\sqrt{1+x}\simeq 1+x/2$. Assuming also that R or z is much greater than the wavelength of the wave, $z\gg\lambda$, so that $k\gg 1/R$ we obtain

$$\frac{2z}{R}\left(jk+\frac{1}{R}\right)\frac{e^{-jkR}}{4\pi R}\simeq jk\frac{e^{-jk(z+|\mathbf{r}_\perp-\mathbf{r}_\perp'|^2/2z)}}{2\pi z}=\frac{jk}{2\pi z}\,e^{-jkz}\,e^{-jk|\mathbf{r}_\perp-\mathbf{r}_\perp'|^2/2z}\qquad(17.18.2)$$

where we set $R\simeq z$ in the amplitude factors, but kept the quadratic approximation in the phase e^{-jkR}. The Fresnel approximation is finally:

$$\boxed{E(\mathbf{r}_\perp,z)=\frac{jk}{2\pi z}\,e^{-jkz}\int_S E(\mathbf{r}_\perp',0)\,e^{-jk|\mathbf{r}_\perp-\mathbf{r}_\perp'|^2/2z}\,d^2\mathbf{r}_\perp'}\quad\text{(Fresnel)}\qquad(17.18.3)$$

This amounts to replacing the propagator impulse response $g(\mathbf{r}_\perp,z)$ by the approximation of Eq. (17.18.2):

$$\boxed{g(\mathbf{r}_\perp,z)=\frac{jk}{2\pi z}\,e^{-jkz}\,e^{-jk|\mathbf{r}_\perp|^2/2z}=\frac{jk}{2\pi z}\,e^{-jkz}\,e^{-jk(x^2+y^2)/2z}}\quad\text{(Fresnel)}\qquad(17.18.4)$$

Noting that $k=2\pi/\lambda$, the constant factor in front is often written as:

$$\frac{jk}{2\pi z}=\frac{j}{\lambda z}$$

The above approximations can also be understood from the plane-wave spectrum point of view. The Fourier transform of (17.18.4) is obtained from the following Fourier integral, which is a special case of (3.5.18):

$$\sqrt{\frac{jk}{2\pi z}}\int_{-\infty}^{\infty}e^{-jkx^2/2z}\,e^{jk_x x}\,dx=e^{jk_x^2z/2k}\qquad(17.18.5)$$

Applying (17.18.5) with respect to the x and y integrations, we obtain the spatial Fourier transform of $g(\mathbf{r}_\perp,z)$:

$$\boxed{\hat g(\mathbf{k}_\perp,z)=e^{-jkz}\,e^{j(k_x^2+k_y^2)z/2k}=e^{-jkz}\,e^{j|\mathbf{k}_\perp|^2z/2k}}\qquad(17.18.6)$$

Then, Eq. (17.18.3) can be written in its plane-wave spectrum form:

$$\boxed{E(\mathbf{r}_\perp,z)=e^{-jkz}\int_S\hat E(\mathbf{k}_\perp,0)\,e^{j|\mathbf{k}_\perp|^2z/2k}\,e^{-j\mathbf{k}_\perp\cdot\mathbf{r}_\perp}\,\frac{d^2\mathbf{k}_\perp}{(2\pi)^2}}\qquad(17.18.7)$$

Eq. (17.18.6) can be obtained from the exact form $\hat g(\mathbf{k}_\perp,z)=e^{-jk_zz}$ by assuming that for large z the evanescent modes will be absent and assuming the approximation $k_x^2+k_y^2\ll k^2$ for the propagating modes. Then, we can write:

$$k_z=\sqrt{k^2-|\mathbf{k}_\perp|^2}=k\sqrt{1-\frac{|\mathbf{k}_\perp|^2}{k^2}}\simeq k\left[1-\frac{1}{2}\frac{|\mathbf{k}_\perp|^2}{k^2}\right]=k-\frac{|\mathbf{k}_\perp|^2}{2k}$$

and, hence

$$e^{-jk_zz}\simeq e^{-jkz}\,e^{j|\mathbf{k}_\perp|^2z/2k}$$

Because of the assumption $|\mathbf{k}_\perp|\ll k$, the maximum transverse wavenumber will be $|\mathbf{k}_\perp|=k=2\pi/\lambda$, and correspondingly the smallest achievable transverse spatial resolution will be $\Delta r_\perp\sim 1/|\mathbf{k}_\perp|\sim\lambda$, that is, about one wavelength. This is the basic diffraction limit of optical instruments, such as lenses and microscopes.

Near-field optics methods [494–512], where the evanescent modes are not ignored, overcome this limitation and can achieve much higher, subwavelength, resolutions.

Although ordinary lenses are diffraction-limited, it has been shown recently [360] that "superlenses" made from metamaterials having negative refractive index can achieve perfect resolution.

In the special case when the aperture field $E(x',y',0)$ depends only on one transverse coordinate, say, $E(x',0)$, the dependence of (17.18.3) on the y direction can be integrated out using the integral

$$\sqrt{\frac{jk}{2\pi z}}\int_{-\infty}^{\infty}e^{-jk(y-y')^2/2z}\,dy'=1\qquad(17.18.8)$$

and we obtain the following one-dimensional version of the Fresnel formula, written convolutionally and in its plane-wave spectrum form:

$$\boxed{\begin{aligned}E(x,z)&=\sqrt{\frac{jk}{2\pi z}}\,e^{-jkz}\int_{-\infty}^{\infty}E(x',0)\,e^{-jk(x-x')^2/2z}\,dx'\\&=e^{-jkz}\int_{-\infty}^{\infty}\hat E(k_x,0)\,e^{jk_x^2z/2k}\,\frac{dk_x}{2\pi}\end{aligned}}\qquad(17.18.9)$$

The *Fraunhofer approximation* is a limiting case of the Fresnel approximation when the distance z is even larger than that in the Fresnel case. More precisely, it is obtained in the far-field limit when $k|\mathbf{r}_\perp'|^2\ll z$, or, $d^2\ll\lambda z$, where d is the size of the aperture.

In this approximation, the field $E(\mathbf{r}_\perp,z)$ becomes proportional to the *Fourier transform* $\hat E(\mathbf{k}_\perp,0)$ of the field at the input plane. It is similar to the radiation-field approximation in which the radiation fields are proportional to Fourier transform of the current sources, that is, to the radiation vector.

A direct way of deriving the Fraunhofer approximation is by applying the stationary-phase approximation—Eq. (F.22) of Appendix F—to the evaluation of the plane-wave spectrum integral (17.18.7). Define the phase function

$$\phi(\mathbf{k}_\perp)=\frac{|\mathbf{k}_\perp|^2z}{2k}-\mathbf{k}_\perp\cdot\mathbf{r}_\perp=\left[\frac{k_x^2z}{2k}-k_xx\right]+\left[\frac{k_y^2z}{2k}-k_yy\right]\equiv\phi_x(k_x)+\phi_y(k_y)$$

Then, the stationary-point with respect to the k_x variable is

$$\phi_x'(k_x)=\frac{k_xz}{k}-x=0\quad\Rightarrow\quad k_x=\frac{xk}{z}\,,\quad\phi_x''(k_x)=\frac{z}{k}$$

and similar expressions for $\phi_y(k_y)$. Thus, vectorially, the stationary point is at $\mathbf{k}_\perp=k\mathbf{r}_\perp/z$. Using Eq. (F.22), we obtain:

$$\int_{-\infty}^{\infty}\hat E(\mathbf{k}_\perp,0)\,e^{j\phi(\mathbf{k}_\perp)}\,\frac{d^2\mathbf{k}_\perp}{(2\pi)^2}\simeq\sqrt{\frac{2\pi j}{\phi_x''(k_x)}\cdot\frac{2\pi j}{\phi_y''(k_y)}}\left[\hat E(\mathbf{k}_\perp,0)\,e^{j\phi(\mathbf{k}_\perp)}\,\frac{1}{(2\pi)^2}\right]_{\mathbf{k}_\perp=\frac{k\mathbf{r}_\perp}{z}}$$

Noting that $\phi(\boldsymbol{k}_\perp) = -k|\boldsymbol{r}_\perp|^2/2z$ at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$, we finally find:

$$\boxed{E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z}\, \big[\hat{E}(\boldsymbol{k}_\perp, 0)\big]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{z}}} \quad \text{(Fraunhofer)} \tag{17.18.10}$$

A simpler way of deriving (17.18.10) is by using (17.18.3) and noting that

$$e^{-jk|\boldsymbol{r}_\perp - \boldsymbol{r}_\perp'|^2/2z} = e^{-jk|\boldsymbol{r}_\perp|^2/2z}\, e^{-jk|\boldsymbol{r}_\perp'|^2/2z}\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}_\perp'/z}$$

The factor $e^{-jk|\boldsymbol{r}_\perp'|^2/2z}$ can be ignored if we assume that $k|\boldsymbol{r}_\perp'|^2 \ll z$, which leads to:

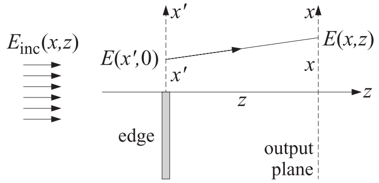
$$E(\boldsymbol{r}_\perp, z) = \frac{jk}{2\pi z}\, e^{-jkz}\, e^{-jk|\boldsymbol{r}_\perp|^2/2z} \int_{-\infty}^{\infty} E(\boldsymbol{r}_\perp', 0)\, e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}_\perp'/z}\, d^2\boldsymbol{r}_\perp'$$

and the last integral factor is recognized as $\hat{E}(\boldsymbol{k}_\perp, 0)$ evaluated at $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/z$.

Example 17.18.1: *Knife-Edge Diffraction.* Let us revisit the problem of knife-edge diffraction using the Fresnel formula (17.18.3). The infinite edge is along the y direction and it occupies the region $x < 0$, as shown in the figure below. The incident plane-wave field and the diffracted field at distance z are:

$$E_{\text{inc}}(x, z) = E_0 e^{-jkz}$$

$$E(x, z) = \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz} \int_{-\infty}^{\infty} E(x', 0)\, e^{-jk(x-x')^2/2z}\, dx'$$



At the input plane, $E(x', 0) = E_0$, for $x' \geq 0$, and $E(x', 0) = 0$, for $x' < 0$. Then, the above integral becomes:

$$E(x, z) = E_0 e^{-jkz} \sqrt{\frac{jk}{2\pi z}} \int_0^{\infty} e^{-jk(x-x')^2/2z}\, dx'$$

Making the change of variables,

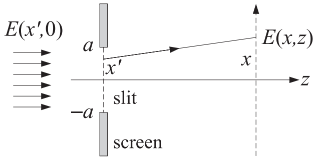
$$\sqrt{\frac{k}{2z}}\,(x' - x) = \sqrt{\frac{\pi}{2}}\, u\,, \qquad v = \sqrt{\frac{k}{\pi z}}\, x$$

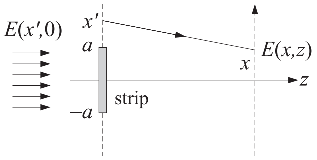
the above integral can be reduced to the Fresnel integral $\mathcal{F}(x)$ of Appendix F:

$$E(x, z) = E_0 e^{-jkz} \sqrt{\frac{j}{2}} \int_{-v}^{\infty} e^{-j\pi u^2/2}\, du = E_0 e^{-jkz} \frac{1}{1-j}\left[\mathcal{F}(v) + \frac{1-j}{2}\right]$$

This is identical (up to the paraxial assumption) to the case discussed in Sec. 17.14. When $x < 0$, the observation point lies in the shadow region. □

Example 17.18.2: *Diffraction by an infinite slit.* Consider an infinite slit on an opaque screen. The y-dimension of the slit is infinite and its x-size is $|x| \leq a$, as shown on the left in the figure below. The same figure also shows an opaque strip of the same size.





The incident field is a uniform plane wave, $E_{\text{inc}}(x, z) = E_0 e^{-jkz}$, whose values on the slit are $E(x', 0) = E_0$. The diffracted field at distance z is given by Eq. (17.18.9):

$$E(x, z) = \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz} \int_{-\infty}^{\infty} E(x', 0)\, e^{-jk(x-x')^2/2z}\, dx' = E_0 \sqrt{\frac{jk}{2\pi z}}\, e^{-jkz} \int_{-a}^{a} e^{-jk(x-x')^2/2z}\, dx'$$

The integral can be reduced to the Fresnel integral $\mathcal{F}(x)$ of Appendix F by making the change of variables:

$$\sqrt{\frac{k}{2z}}\,(x' - x) = \sqrt{\frac{\pi}{2}}\, u\,, \quad v_\pm = \sqrt{\frac{k}{\pi z}}\,(\pm a - x)$$

so that

$$\sqrt{\frac{jk}{2\pi z}} \int_{-a}^{a} e^{-jk(x-x')^2/2z}\, dx' = \sqrt{\frac{j}{2}} \int_{v_-}^{v_+} e^{-j\pi u^2/2}\, du = \frac{\mathcal{F}(v_+) - \mathcal{F}(v_-)}{1-j} \equiv D(x, z)$$

where we used $\sqrt{j/2} = 1/(1-j)$. Thus, $E(x, z)$ becomes:

$$E(x, z) = e^{-jkz} D(x, z) \tag{17.18.11}$$

For the case of the strip, the limits of integration are changed to:

$$\sqrt{\frac{jk}{2\pi z}} \left(\int_a^{\infty} + \int_{-\infty}^{-a}\right) e^{-jk(x-x')^2/2z}\, dx' = \frac{\mathcal{F}(\infty) - \mathcal{F}(v_+) + \mathcal{F}(v_-) - \mathcal{F}(-\infty)}{1-j} = 1 - D(x, z)$$

where we used $\mathcal{F}(\infty) = -\mathcal{F}(-\infty) = (1-j)/2$. Thus, the diffracted field in the strip case will be given by the complementary expression

$$E(x, z) = e^{-jkz}\big[1 - D(x, z)\big] \tag{17.18.12}$$

This result is an example of the Babinet principle [598] that the sum of the fields from an aperture and its complementary screen is equal to the field in the absence of the aperture:

$$E_{\text{slit}}(x, z) + E_{\text{strip}}(x, z) = e^{-jkz}$$

Fig. 17.18.1 shows the diffracted patterns in the two cases. The graphs plot the quantities $|D(x, z)|$ and $|1 - D(x, z)|$ versus x in the two cases.

The slit was chosen to be four wavelengths wide, $a = 4\lambda$, and the diffracted patterns correspond to the near, medium, and far distances $z = a$, $z = 20a$, and $z = 100a$. The latter case corresponds to the Fraunhofer pattern having a small ratio $a^2/\lambda z = 1/25$. For example, for the slit case, the corresponding pattern approximates (but it is not quite there yet) the typical sinc-function Fourier transform of the rectangular slit distribution $E(x', 0) = E_0$, for $-a \leq x' \leq a$:

$$\hat{E}(k_x, 0) = \int_{-a}^{a} E_0\, e^{jk_x x'}\, dx' = 2aE_0\, \frac{\sin(k_x a)}{k_x a}$$

where this is to be evaluated at $k_x = kx/z$ for the diffraction pattern $E(x, z)$. The property that at the center of the strip, $x = 0$, the diffracted pattern is not zero is an example of the so-called *Poisson's spot* [598]. □

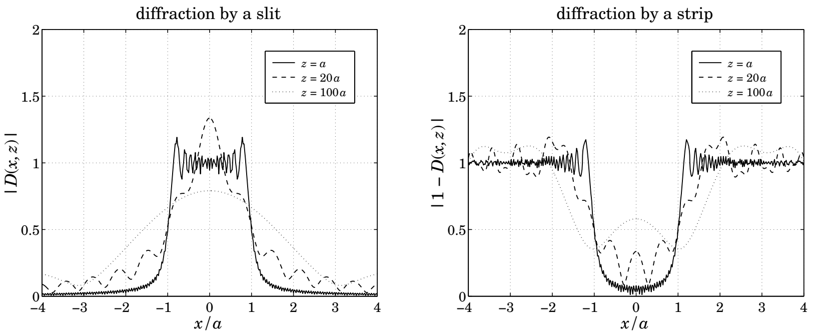


Fig. 17.18.1 Fresnel diffraction by a slit and a strip

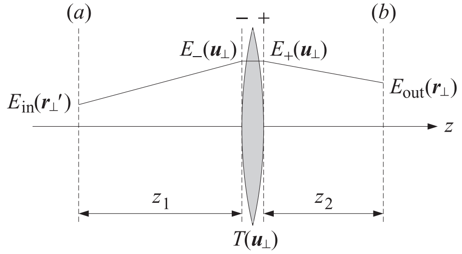

(a)
(b)
− +
$E_-(\boldsymbol{u}_\perp)$
$E_+(\boldsymbol{u}_\perp)$
$E_{\rm in}(\boldsymbol{r}_\perp')$
$E_{\rm out}(\boldsymbol{r}_\perp)$
z
z_1
z_2
$T(\boldsymbol{u}_\perp)$

Fig. 17.18.2 Field propagated from plane *a* to plane *b* through a thin optical element.

In Fourier optics applications, one considers the passage of light through various optical elements that perform certain functions, such as Fourier transformation using lenses. For example, Fig. 17.18.2 shows an input field starting at aperture plane *a*, then propagating a distance z_1 to a thin optical element where it is modified by a transmittance function, and then propagating another distance z_2 to an aperture plane *b*.

Assuming that the input/output relationship of the optical element is multiplicative, $E_+(\boldsymbol{r}_\perp)= T(\boldsymbol{r}_\perp)E_-(\boldsymbol{r}_\perp)$, the relationship between the output field at plane *b* to the input field at plane *a* is obtained by successively applying the propagation equation (17.17.10):

$$\begin{aligned}
E_{\rm out}(\boldsymbol{r}_\perp) &= \int_S g(\boldsymbol{r}_\perp-\boldsymbol{u}_\perp, z_2)E_+(\boldsymbol{u}_\perp)\,d^2\boldsymbol{u}_\perp = \int_S g(\boldsymbol{r}_\perp-\boldsymbol{u}_\perp, z_2)T(\boldsymbol{u}_\perp)E_-(\boldsymbol{u}_\perp)\,d^2\boldsymbol{u}_\perp\\
&= \int_S g(\boldsymbol{r}_\perp-\boldsymbol{u}_\perp, z_2)T(\boldsymbol{u}_\perp)g(\boldsymbol{u}_\perp-\boldsymbol{r}_\perp', z_1)E_{\rm in}(\boldsymbol{r}_\perp')\,d^2\boldsymbol{u}_\perp\,d^2\boldsymbol{r}_\perp'\\
&= \int_S h(\boldsymbol{r}_\perp,\boldsymbol{r}_\perp')E_{\rm in}(\boldsymbol{r}_\perp')\,d^2\boldsymbol{r}_\perp'
\end{aligned}$$

where the overall transfer function from plane *a* to plane *b* will be:

$$h(\boldsymbol{r}_\perp,\boldsymbol{r}_\perp')= \int_S g(\boldsymbol{r}_\perp-\boldsymbol{u}_\perp, z_2)T(\boldsymbol{u}_\perp)g(\boldsymbol{u}_\perp-\boldsymbol{r}_\perp', z_1)\,d^2\boldsymbol{u}_\perp \tag{17.18.13}$$

where we labeled the spatial x, y coordinates by $\boldsymbol{r}_\perp', \boldsymbol{u}_\perp$, and $\boldsymbol{r}_\perp$ on the planes (a), the element, and plane (b).

In a similar fashion, one can work out the transfer function of more complicated configurations. For example, passing through two transmittance elements as shown in Fig. 17.18.3, we will have:

$$E_{\rm out}(\boldsymbol{r}_\perp)= \int_S h(\boldsymbol{r}_\perp,\boldsymbol{r}_\perp')E_{\rm in}(\boldsymbol{r}_\perp')\,d^2\boldsymbol{r}_\perp' \tag{17.18.14}$$

where

$$h(\boldsymbol{r}_\perp,\boldsymbol{r}_\perp')= \int_S g(\boldsymbol{r}_\perp-\boldsymbol{u}_\perp, z_2)T_2(\boldsymbol{u}_\perp)g(\boldsymbol{u}_\perp-\boldsymbol{v}_\perp, z_0)T_1(\boldsymbol{v}_\perp)g(\boldsymbol{v}_\perp-\boldsymbol{r}_\perp', z_1)\,d^2\boldsymbol{u}_\perp\,d^2\boldsymbol{v}_\perp \tag{17.18.15}$$

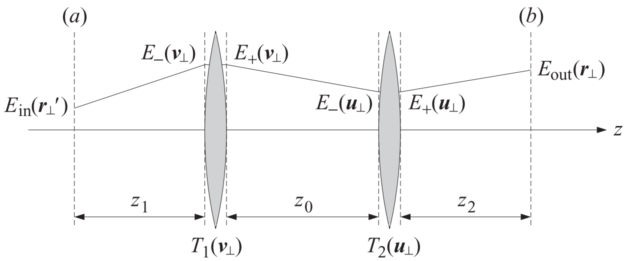

(a)
(b)
$E_-(\boldsymbol{v}_\perp)$
$E_+(\boldsymbol{v}_\perp)$
$E_{\rm out}(\boldsymbol{r}_\perp)$
$E_{\rm in}(\boldsymbol{r}_\perp')$
$E_-(\boldsymbol{u}_\perp)$
$E_+(\boldsymbol{u}_\perp)$
z
z_1
z_0
z_2
$T_1(\boldsymbol{v}_\perp)$
$T_2(\boldsymbol{u}_\perp)$

Fig. 17.18.3 Field propagated from plane *a* to plane *b* through multiple optical elements.

Lenses are probably the most important optical elements. Their interesting properties arise from their transmittance function, which has the quadratic phase:

$$\boxed{T(\boldsymbol{r}_\perp)= e^{jk|\boldsymbol{r}_\perp|^2/2F} = e^{jk(x^2+y^2)/2F}} \quad \text{(lens transmittance)} \tag{17.18.16}$$

where F is the focal length. Because the Fresnel propagation factor $e^{-jk|\boldsymbol{r}_\perp|^2/2z}$ also has the same type of quadratic phase, but with the *opposite* sign, it is possible for lenses to act as spatial "dispersion compensation" elements, much like the dispersion compensation and pulse compression filters of Chap. 3. They have many uses, such as compensating propagation effects and focusing the waves on appropriate planes, or performing spatial Fourier transforms.

The transmittance function (17.18.16) can be derived with the help of Fig. 17.18.4, which shows a wave entering from the left a (convex) spherical glass surface at a distance x from the axis.

Let R and d denote the radius of the spherical element, and its maximum width along its axis to the flat back plane, and let n be its refractive index. The wave travels a distance a in air and a distance b in the glass. If k is the free-space wavenumber, then in the glass it changes to $k_g = kn$. Therefore, the wave will accumulate the following phase as it propagates from the front plane to the back plane:

$$e^{-jka}\,e^{-jk_g b}$$

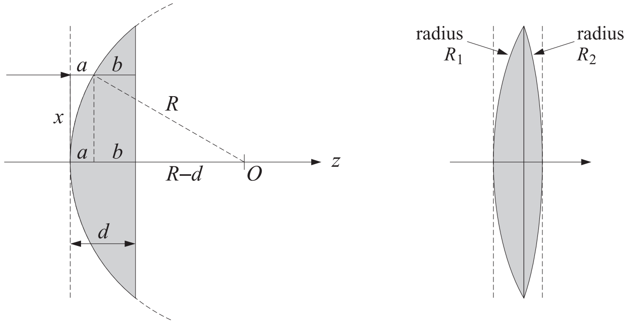


Fig. 17.18.4 Transmittance of a thin spherical lens.

where we are assuming a thin lens, which allows us to ignore the bending of the ray arising from refraction. Because, $a + b = d$, we have for the net phase:

$$\phi(x) = ka + k_g b = ka + nk(d - a) = nkd - (n-1)ka$$

The distance a is easily seen from the above figure to be:

$$a = R - \sqrt{R^2 - x^2}$$

Assuming that $x \ll R$, we can expand the square root to get:

$$a = R - R\sqrt{1 - \frac{x^2}{R^2}} \simeq R - R\left[1 - \frac{1}{2}\frac{x^2}{R^2}\right] = \frac{x^2}{2R}$$

Thus, the phase $\phi(x)$ is approximately,

$$\phi(x) = knd - (n-1)ka = knd - \frac{(n-1)kx^2}{2R}$$

If we make up a convex lens by putting together two such spherical lenses with radii R_1 and R_2, as shown in Fig. 17.18.4, then the net phase change between the front and back planes will be, ignoring the constant nkd terms:

$$\phi(x) = -(n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)\frac{kx^2}{2} \equiv -\frac{kx^2}{2F} \tag{17.18.17}$$

where we defined the focal length F of the lens through the "lensmaker's equation,"

$$\frac{1}{F} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{17.18.18}$$

In a two-dimensional description, we replace x^2 by $|\mathbf{r}_\perp|^2 = x^2 + y^2$. Thus, the phase change and corresponding transmittance function will be:

$$\phi(\mathbf{r}_\perp) = -\frac{k|\mathbf{r}_\perp|^2}{2F} \quad \Rightarrow \quad T(\mathbf{r}_\perp) = e^{-j\phi(\mathbf{r}_\perp)} = e^{jk|\mathbf{r}_\perp|^2/2F}$$

Some examples of the various effects that can be accomplished with lenses can be obtained by applying the configurations of Figs. 17.18.2 and 17.18.3 with appropriate choices of the aperture planes and focal lengths. We will use the Fresnel approximation (17.18.4) for $g(\mathbf{r}_\perp, z)$ in all of the examples and assume that the transmittance (17.18.16) extends over the entire xy plane—in effect, we are replacing the lens with the ideal case of an infinitely thin transparency with transmittance (17.18.16).

The main property of a lens is to turn an incident plane wave from the left into a spherical wave converging on the lens focus on the right, and similarly, if a source of a spherical wave is placed at the focus on the left, then the diverging wave will be converted into a plane wave after it passes through the lens. These cases are shown in Fig. 17.18.5.

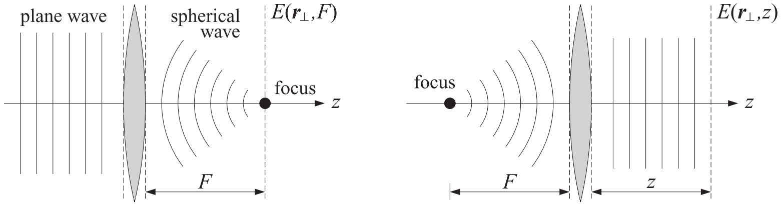


Fig. 17.18.5 Spherical waves converging to, or diverging from, a lens focal point.

The case on the left corresponds to the choices $z_1 = 0$ and $z_2 = F$ in Fig. 17.18.2, that is, the input plane coincides with the left plane of the lens. The incident wave has a constant amplitude on the plane $E_{\text{in}}(\mathbf{r}_\perp) = E_0$. Noting that $g(\mathbf{r}'_\perp - \mathbf{u}_\perp, 0) = \delta^2(\mathbf{r}'_\perp - \mathbf{u}_\perp)$, we obtain from Eq. (17.18.13) with $z_2 = F$:

$$h(\mathbf{r}_\perp, \mathbf{r}'_\perp) = T(\mathbf{r}'_\perp)g(\mathbf{r}_\perp - \mathbf{r}'_\perp, F) = \frac{jk}{2\pi F}\, e^{-jkF}\, e^{jk|\mathbf{r}'_\perp|^2/2F}\, e^{-jk|\mathbf{r}_\perp - \mathbf{r}'_\perp|^2/2F}$$

the quadratic phase terms combine as follows:

$$e^{jk|\mathbf{r}'_\perp|^2/2F}\, e^{-jk|\mathbf{r}_\perp - \mathbf{r}'_\perp|^2/2F} = e^{-jk|\mathbf{r}_\perp|^2/2F}\, e^{jk\mathbf{r}_\perp\cdot\mathbf{r}'_\perp/F}$$

and result in the following transfer function:

$$h(\mathbf{r}_\perp, \mathbf{r}'_\perp) = T(\mathbf{r}'_\perp)g(\mathbf{r}_\perp - \mathbf{r}'_\perp, F) = \frac{jk}{2\pi F}\, e^{-jkF}\, e^{-jk|\mathbf{r}_\perp|^2/2F}\, e^{jk\mathbf{r}_\perp\cdot\mathbf{r}'_\perp/F} \tag{17.18.19}$$

Its integration with the constant input results in:

$$E(\mathbf{r}_\perp, F) = \frac{jk}{2\pi F}\, e^{-jkF}\, E_0 e^{-jk|\mathbf{r}_\perp|^2/2F} \int_{-\infty}^{\infty} e^{jk\mathbf{r}_\perp\cdot\mathbf{r}'_\perp/F}\, d^2\mathbf{r}'_\perp$$

The integral is equal to the Dirac delta, $(2\pi)^2\delta(k\mathbf{r}_\perp/F) = (2\pi)^2\delta(\mathbf{r}_\perp)F^2/k^2$. Thus,

$$E(\mathbf{r}_\perp, F) = \frac{j2\pi F}{k}\, e^{-jkF}\, E_0\, \delta(\mathbf{r}_\perp)$$

which is sharply focused onto the focal point $\boldsymbol{r}_\perp = 0$ and $z = F$. For the second case depicted on the right in Fig. 17.18.5, we first note that the paraxial approximation for a spherical wave placed at the origin is:

$$E_0\,\frac{e^{-jkr}}{4\pi r} \simeq E_0\,\frac{jk}{4\pi z}\,e^{-jkz}\,e^{-jk|\boldsymbol{r}_\perp|^2/2z}\,,\quad r \simeq z - \frac{|\boldsymbol{r}_\perp|^2}{2z}$$

If this source is placed at the left focal point of the lens, then, the diverging paraxial spherical wave, after traveling distance $z = F$, will arrive at the left plane of the lens:

$$E_{\rm in}(\boldsymbol{r}'_\perp, 0) = E_0\,\frac{jk}{4\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}'_\perp|^2/2F} \equiv E_1\,e^{-jk|\boldsymbol{r}'_\perp|^2/2F}$$

The transmittance of the lens will compensate this propagation phase resulting into a constant field at the output plane of the lens, which will then propagate to the right as a plane wave:

$$E(\boldsymbol{r}'_\perp, 0) = T(\boldsymbol{r}'_\perp)E_{\rm in}(\boldsymbol{r}'_\perp, 0) = e^{jk|\boldsymbol{r}'_\perp|^2/2F}\,E_1\,e^{-jk|\boldsymbol{r}'_\perp|^2/2F} = E_1$$

The propagated field to distance z is obtained from Eq. (17.18.3):

$$E(\boldsymbol{r}_\perp, z) = e^{-jkz}\,\frac{jk}{2\pi z}\int_{-\infty}^{\infty} E_1 e^{-jk|\boldsymbol{r}_\perp-\boldsymbol{r}'_\perp|^2/2z}\,d^2\boldsymbol{r}'_\perp = E_1\,e^{-jkz}\,\frac{jk}{2\pi z}\,\frac{2\pi z}{jk} = E_1 e^{-jkz}$$

where the integral was evaluated using twice the result (17.18.8). Thus, the transmitted wave is a uniform plane wave propagating along the z-direction.

To see the Fourier transformation property of lenses, consider again the left picture in Fig. 17.18.5 with the output plane still placed at the right focal length $z_2 = F$, but take an arbitrary field $E_{\rm in}(\boldsymbol{r}'_\perp)$ incident at the left plane of the lens. The overall transfer function is still the same as in Eq. (17.18.19), thus, giving:

$$\begin{aligned} E(\boldsymbol{r}_\perp, F) &= \int_{-\infty}^{\infty} h(\boldsymbol{r}_\perp - \boldsymbol{r}'_\perp)E_{\rm in}(\boldsymbol{r}'_\perp)\,d^2\boldsymbol{r}'_\perp \\ &= \frac{jk}{2\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}_\perp|^2/2F}\int_{-\infty}^{\infty} E_{\rm in}(\boldsymbol{r}'_\perp)\,e^{jk\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp/F}\,d^2\boldsymbol{r}'_\perp \end{aligned} \tag{17.18.20}$$

The last integral factor is recognized as the Fourier transform $\hat{E}_{\rm in}(\boldsymbol{k}_\perp)$ evaluated at wavenumber $\boldsymbol{k}_\perp = k\boldsymbol{r}_\perp/F$. Thus, we obtain:

$$E(\boldsymbol{r}_\perp, F) = \frac{jk}{2\pi F}\,e^{-jkF}\,e^{-jk|\boldsymbol{r}_\perp|^2/2F}\big[\hat{E}_{\rm in}(\boldsymbol{k}_\perp)\big]_{\boldsymbol{k}_\perp = \frac{k\boldsymbol{r}_\perp}{F}} \tag{17.18.21}$$

This result is similar to the Fraunhofer case (17.18.10), but it is valid at the much shorter Fresnel distance $z = F$, instead of the far-field distances. It is analogous to the output of the pulse compression filter in chirp radar discussed in Chap. 3, see for example Eq. (3.10.14).

It is left as an exercise to show that the extra quadratic phase factor in (17.18.21) can be eliminated by using the configuration of Fig. 17.18.2 with both aperture planes placed at the foci of the lens, that is, $z_1 = z_2 = F$, (known as a $2F$ system.)

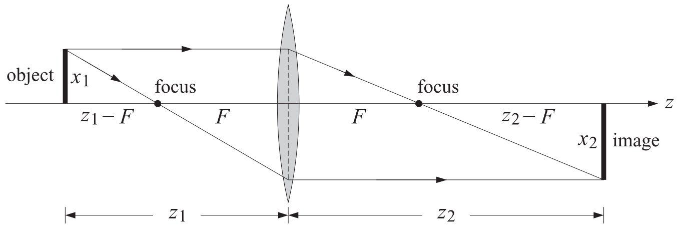


Fig. 17.18.6 Lens law of magnification.

Finally, let us look at the magnifying properties of a lens. Fig. 17.18.6 shows an image placed at distance z_1 from the left and its image formed at distance z_2 on the right. It is well-known that the distances z_1, z_2 must be related by:

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{F} \tag{17.18.22}$$

The magnification law relates the size of the image to the size of the object:

$$M = \frac{x_2}{x_1} = \frac{z_2}{z_1} \qquad \text{(magnification ratio)} \tag{17.18.23}$$

These properties can be derived by tracing the rays emanating from the top of the object. The ray that is parallel to the lens axis will bend to pass through the focal point on the right. The ray from the top of the object through the left focal point will bend to become parallel to the axis. The intersection of these two rays defines the top point of the image. From the geometry of the graph one has:

$$\frac{x_1}{z_1 - F} = \frac{x_2}{F} \quad \text{and} \quad \frac{x_2}{z_2 - F} = \frac{x_1}{F}$$

The consistency of the equations requires the condition $(z_1 - F)(z_2 - F) = F^2$, which is equivalent to (17.18.22). Then, Eq. (17.18.23) follows by replacing F from (17.18.22) into the ratio $x_2/x_1 = (z_2 - F)/F$.

To understand (17.18.22) and (17.18.23) from the point of view of Fresnel diffraction, we note that the transfer function (17.18.13) involves the following quadratic phase factors, with the middle one being the lens transmittance:

$$\begin{aligned} &e^{-jk|\boldsymbol{r}_\perp-\boldsymbol{u}_\perp|^2/2z_2}\,e^{jk|\boldsymbol{u}_\perp|^2/2F}\,e^{-jk|\boldsymbol{u}_\perp-\boldsymbol{r}'_\perp|^2/2z_1} \\ &= e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}\,e^{-jk|\boldsymbol{r}'_\perp|^2/2z_1}\,e^{-jk(1/z_1+1/z_2-1/F)|\boldsymbol{u}_\perp|^2/2}\,e^{jk\boldsymbol{u}_\perp\cdot(\boldsymbol{r}_\perp/z_2+\boldsymbol{r}'_\perp/z_1)} \end{aligned}$$

Because of Eq. (17.18.22), the term that depends quadratically on $\boldsymbol{u}_\perp$ cancels and one is left only with a linear dependence on $\boldsymbol{u}_\perp$. This integrates into a delta function in (17.18.13), resulting in

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \frac{jke^{-jkz_1}}{2\pi z_1}\,\frac{jke^{-jkz_2}}{2\pi z_2}\,e^{-jk|\boldsymbol{r}_\perp|^2/2z_2}\,e^{-jk|\boldsymbol{r}'_\perp|^2/2z_1}\,(2\pi)^2\delta\left(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}'_\perp}{z_1}\right)$$

The delta function forces $\boldsymbol{r}_\perp = -(z_2/z_1)\boldsymbol{r}_\perp'$, which is the same as (17.18.23). The negative sign means that the image is upside down. Noting that

$$\delta\left(\frac{k\boldsymbol{r}_\perp}{z_2} + \frac{k\boldsymbol{r}_\perp'}{z_1}\right) = \frac{z_1^2}{k^2}\,\delta\left(\boldsymbol{r}_\perp' + \frac{z_1}{z_2}\,\boldsymbol{r}_\perp\right)$$

we obtain for the field at the output plane:

$$E_{\text{out}}(\boldsymbol{r}_\perp) = -\frac{z_1}{z_2}\,E_{\text{in}}\left(-\frac{z_1}{z_2}\,\boldsymbol{r}_\perp\right)\,e^{-jk(z_1+z_2)}\,e^{-jk|\boldsymbol{r}_\perp|^2(z_1+z_2)/2z_2^2}$$

which represents a scaled and reversed version of the input.

Some references on the Rayleigh-Sommerfeld diffraction theory, the plane-wave spectrum representation, and Fourier optics are [1110,1111] and [1148–1161].

17.19 Problems

17.1 Show that Eq. (17.4.9) can be written in the compact vectorial form:

$$\boldsymbol{E} = -jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\left[\hat{\mathbf{z}}\times\boldsymbol{f} - \eta\,\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{g})\right],\quad \boldsymbol{H} = -\frac{jk}{\eta}\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\left[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{f}) + \eta\,\hat{\mathbf{z}}\times\boldsymbol{g}\right]$$

Similarly, show that Eqs. (17.4.10) and (17.4.11) can be written as:

$$\boldsymbol{E} = -2jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{z}}\times\boldsymbol{f}],\qquad \boldsymbol{H} = -\frac{2jk}{\eta}\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{f})]$$

$$\boldsymbol{E} = 2jk\eta\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{r}}\times(\hat{\mathbf{z}}\times\boldsymbol{g})],\quad \boldsymbol{H} = -2jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times[\hat{\mathbf{z}}\times\boldsymbol{g}]$$

17.2 Prove the first pair of equations for $\boldsymbol{E},\boldsymbol{H}$ of the previous problem by working exclusively with the Kottler formulas (17.4.2) and taking their far-field limits.

17.3 Explain in detail how the inequality (17.6.12) for the aperture efficiency e_a may be thought of as an example of the Schwarz inequality. Then, using standard properties of Schwarz inequalities, prove that the maximum of e_a is unity and is achieved for uniform apertures. As a reminder, the Schwarz inequality for single-variable complex-valued functions is:

$$\left|\int_a^b f^*(x)\,g(x)\,dx\right|^2 \le \int_a^b |f(x)|^2\,dx\cdot\int_a^b |g(x)|^2\,dx$$

17.4 To prove the equivalence of the Kirchhoff diffraction and Stratton-Chu formulas, (17.10.6) and (17.10.7), use the identities (C.29) and (C.32) of Appendix C, to obtain:

$$\int_V\left[j\omega\mu\,G\,\boldsymbol{J} + \frac{1}{\epsilon}G\,\nabla'\rho + G\,\nabla'\times\boldsymbol{J}_m\right]dV' = \int_V\left[j\omega\mu\,G\,\boldsymbol{J} - \frac{\rho}{\epsilon}\,\nabla'G + \boldsymbol{J}_m\times\nabla'G\right]dV'$$
$$-\oint_S\left[\hat{\mathbf{n}}\,\frac{\rho}{\epsilon}\,G + \hat{\mathbf{n}}\times\boldsymbol{J}_m G\right]dS'$$

Then, using the identity (C.33), show that Eq. (17.10.6) can be rewritten in the form:

$$\begin{aligned}\boldsymbol{E}(\mathbf{r}) = &-\int_V\left[j\omega\mu\,G\,\boldsymbol{J} - \frac{\rho}{\epsilon}\,\nabla'G + \boldsymbol{J}_m\times\nabla'G\right]dV'\\ &+\oint_S\left[\hat{\mathbf{n}}\,\frac{\rho}{\epsilon}\,G + \hat{\mathbf{n}}\times\boldsymbol{J}_m G\right]dS'\\ &-\oint_S\left[\hat{\mathbf{n}}\,G\,\nabla'\cdot\boldsymbol{E} - (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla'G - G\,\hat{\mathbf{n}}\times(\nabla'\times\boldsymbol{E}) - (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla'G\right]dS'\end{aligned}$$

Finally, use $\rho/\epsilon = \nabla'\cdot\boldsymbol{E}$ and $\nabla'\times\boldsymbol{E} + \boldsymbol{J}_m = -j\omega\mu\boldsymbol{H}$ to obtain (17.10.7).

17.5 Prove the equivalence of the Stratton-Chu and Kottler formulas, (17.10.7) and (17.10.10), by first proving and then using the following dual relationships:

$$\int_V[j\omega\rho\nabla'G - (\boldsymbol{J}\cdot\nabla')\nabla'G]dV' = \oint_S[((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla')\nabla'G - j\omega\epsilon(\hat{\mathbf{n}}\cdot\boldsymbol{E})\nabla'G]$$

$$\int_V[j\omega\rho_m\nabla'G - (\boldsymbol{J}_m\cdot\nabla')\nabla'G]dV' = -\oint_S[((\hat{\mathbf{n}}\times\boldsymbol{E})\cdot\nabla')\nabla'G + j\omega\mu(\hat{\mathbf{n}}\cdot\boldsymbol{H})\nabla'G]$$

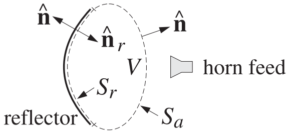
To prove these, work component-wise, use Maxwell's equations (17.2.1), and apply the divergence theorem on the volume V of Fig. 17.10.1.

17.6 Prove the equivalence of the Kottler and Franz formulas, (17.10.10) and (17.10.11), by using the identity $\nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A}$, and by replacing the quantity $k^2G(\mathbf{r}-\mathbf{r}')$ by $-\delta^{(3)}(\mathbf{r}-\mathbf{r}') - \nabla'^2 G$. Argue that the term $\delta^{(3)}(\mathbf{r}-\mathbf{r}')$ makes a difference only for the volume integrals, but not for the surface integrals.

17.7 Prove the equivalence of the modified Stratton-Chu and Kirchhoff diffraction integral formulas of Eq. (17.12.1) and (17.12.2) by using the identity (C.42) of Appendix C and replacing $\nabla'\cdot\boldsymbol{E} = 0$ and $\nabla'\times\boldsymbol{E} = -j\omega\mu\boldsymbol{H}$ in the source-less region under consideration.

17.8 Prove the equivalence of the Kottler and modified Stratton-Chu formulas of Eq. (17.12.1) and (17.12.2) by subtracting the two expressions, replacing $j\omega\epsilon\boldsymbol{E} = \nabla'\times\boldsymbol{H}$, and using the Stokes identity (C.38) of Appendix C.

17.9 Consider a reflector antenna fed by a horn, as shown on the right. A closed surface $S = S_r + S_a$ is such that the portion S_r caps the reflector and the portion S_a is an aperture in front of the reflector. The feed lies outside the closed surface, so that the volume V enclosed by S is free of current sources.



Applying the Kottler version of the extinction theorem of Sec. 17.10 on the volume V, show that for points $\mathbf{r}$ outside V, the field radiated by the induced surface currents on the reflector S_r is equal to the field radiated by the aperture fields on S_a, that is,

$$\begin{aligned}\boldsymbol{E}_{\text{rad}}(\mathbf{r}) &= \frac{1}{j\omega\epsilon}\int_{S_r}[k^2G\,\boldsymbol{J}_s + (\boldsymbol{J}_s\cdot\nabla')\nabla'G]\,dS'\\ &= \frac{1}{j\omega\epsilon}\int_{S_a}[k^2G(\hat{\mathbf{n}}\times\boldsymbol{H}) + ((\hat{\mathbf{n}}\times\boldsymbol{H})\cdot\nabla')\nabla'G + j\omega\epsilon(\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla'G]\,dS'\end{aligned}$$

where the induced surface currents on the reflector are $\boldsymbol{J}_s = \hat{\mathbf{n}}_r\times\boldsymbol{H}$ and $\boldsymbol{J}_{ms} = -\hat{\mathbf{n}}_r\times\boldsymbol{E}$, and on the perfectly conducting reflector surface, we must have $\boldsymbol{J}_{ms} = 0$.

This result establishes the equivalence of the so-called aperture-field and current-distribution methods for reflector antennas [1168].

17.10 Consider an x-polarized uniform plane wave incident obliquely on the straight-edge aperture of Fig. 17.14.4, with a wave vector direction $\hat{\mathbf{k}}_1 = \hat{\mathbf{z}}\cos\theta_1 + \hat{\mathbf{y}}\sin\theta_1$. First show that the tangential fields at an aperture point $\mathbf{r}' = x'\,\hat{\mathbf{x}} + y'\,\hat{\mathbf{y}}$ on the aperture above the straight-edge are given by:

$$\boldsymbol{E}_a = \hat{\mathbf{x}}\,E_0 e^{-jky'\sin\theta_1},\qquad \boldsymbol{H}_a = \hat{\mathbf{y}}\,\frac{E_0}{\eta_0}\,\cos\theta_1 e^{-jky'\sin\theta_1}$$

Then, using Kottler's formula (17.12.1), and applying the usual Fresnel approximations in the integrand, as was done for the point source in Fig. 17.14.4, show that the diffracted wave below the edge is given by Eqs. (17.14.22)–(17.14.24), except that the field at the edge is $E_{\text{edge}} = E_0$, and the focal lengths are in this case $F = l_2$ and $F' = l_2/\cos^2\theta_2$

Finally, show that the asymptotic diffracted field (when $l_2 \to \infty$), is given near the forward direction $\theta \simeq 0$ by:

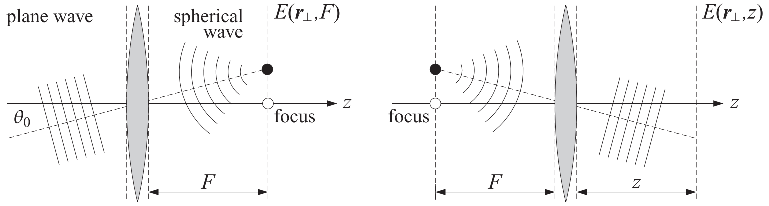
$$E = E_{\text{edge}}\,\frac{e^{-jkl_2}}{\sqrt{l_2}}\,\frac{1-j}{2\sqrt{\pi k}\,\theta}$$

17.11 Assume that the edge in the previous problem is a perfectly conducting screen. Using the field-equivalence principle with effective current densities on the aperture above the edge $\boldsymbol{J}_s = 0$ and $\boldsymbol{J}_{ms} = -2\hat{\mathbf{n}}\times\boldsymbol{E}_a$, and applying the usual Fresnel approximations, show that the diffracted field calculated by Eq. (17.4.1) is is still given by Eqs. (17.14.22)–(17.14.24), except that the factor $\cos\theta_1 + \cos\theta_2$ is replaced now by $2\cos\theta_2$, and that the asymptotic field and edge-diffraction coefficient are:

$$E = E_0\,\frac{e^{-jkl_2}}{\sqrt{l_2}}\,D_{\text{edge}}\,,\qquad D_{\text{edge}} = \frac{(1-j)\,2\cos\theta_2}{4\sqrt{\pi k}\,(\sin\theta_1 + \sin\theta_2)}$$

Show that this expression agrees with the exact Sommerfeld solution (17.15.26) at normal incidence and near the forward diffracted direction.

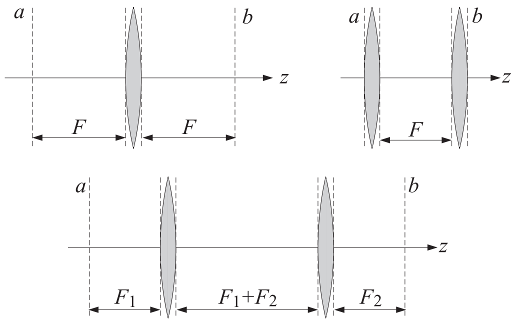
17.12 A uniform plane wave, $E(x,z) = E_0 e^{-jk(x\sin\theta_0 + z\cos\theta_0)}$, is incident obliquely on a lens at an angle θ_0 with the z axis, as shown in the figure below.



Using similar methods as for Fig. 17.18.5, show that after passing through the lens, the wave will converge onto the shifted focal point with coordinates $z = F$ and $x = F\sin\theta_0$.

Conversely, consider a point source of a spherical wave starting at the point $z = -F$ and $x = F\sin\theta_0$. Show that upon passage through the lens, the spherical wave will be converted into the obliquely moving plane wave $E(x,z) = E_1 e^{-jk(x\sin\theta_0 + z\cos\theta_0)}$. What is E_1?

17.13 Consider the three lens configurations shown below. They are special cases of Figs. 17.18.2 and 17.18.3, with appropriate choices for the input and output aperture planes a and b.



Working with Eqs. (17.18.13) and (17.18.15), show that the transfer functions $h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp)$ are given as follows for the three cases:

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = e^{-2jkF} e^{jk(\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp)/F}\,,\qquad h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = \frac{jk}{2\pi F}\,e^{-jkF} e^{jk(\boldsymbol{r}_\perp\cdot\boldsymbol{r}'_\perp)/F}$$

$$h(\boldsymbol{r}_\perp, \boldsymbol{r}'_\perp) = -\frac{F_1}{F_2}\,\delta\left(\boldsymbol{r}_\perp + \frac{F_1}{F_2}\,\boldsymbol{r}_\perp\right)$$

Show that the first two cases perform a Fourier transformation as in Eq. (17.18.21), but without the quadratic phase factors. Show that the third case, performs a scaling of the input with a magnification factor $M = -F_2/F_1$

18

Aperture Antennas

18.1 Open-Ended Waveguides

The aperture fields over an open-ended waveguide are not uniform over the aperture. The standard assumption is that they are equal to the fields that would exist if the guide were to be continued [1].

Fig. 18.1.1 shows a waveguide aperture of dimensions $a > b$. Putting the origin in the middle of the aperture, we assume that the tangential aperture fields $\mathbf{E}_a$, $\mathbf{H}_a$ are equal to those of the TE_{10} mode. We have from Eq. (9.4.3):

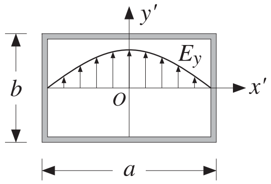


Fig. 18.1.1 Electric field over a waveguide aperture.

$$E_y(x') = E_0 \cos\left(\frac{\pi x'}{a}\right), \quad H_x(x') = -\frac{1}{\eta_{TE}} E_0 \cos\left(\frac{\pi x'}{a}\right) \tag{18.1.1}$$

where $\eta_{TE} = \eta/K$ with $K = \sqrt{1 - \omega_c^2/\omega^2} = \sqrt{1 - (\lambda/2a)^2}$. Note that the boundary conditions are satisfied at the left and right walls, $x' = \pm a/2$.

For larger apertures, such as $a > 2\lambda$, we may set $K \simeq 1$. For smaller apertures, such as $0.5\lambda \le a \le 2\lambda$, we will work with the generalized Huygens source condition (17.5.7). The radiated fields are given by Eq. (17.5.5), with $f_x = 0$:

$$\begin{aligned} E_\theta &= jk\frac{e^{-jkr}}{2\pi r}\, c_\theta f_y(\theta,\phi)\sin\phi \\ E_\phi &= jk\frac{e^{-jkr}}{2\pi r}\, c_\phi f_y(\theta,\phi)\cos\phi \end{aligned} \tag{18.1.2}$$

where $f_y(\theta,\phi)$ is the aperture Fourier transform of $E_y(x')$, that is,

$$\begin{aligned} f_y(\theta,\phi) &= \int_{-a/2}^{a/2}\int_{-b/2}^{b/2} E_y(x')\,e^{jk_x x'+jk_y y'}\,dx'dy' \\ &= E_0\int_{-a/2}^{a/2}\cos\left(\frac{\pi x'}{a}\right)e^{jk_x x'}\,dx'\cdot\int_{-b/2}^{b/2} e^{jk_y y'}\,dy' \end{aligned}$$

The y'-integration is the same as that for a uniform line aperture. For the x'-integration, we use the definite integral:

$$\int_{-a/2}^{a/2}\cos\left(\frac{\pi x'}{a}\right)e^{jk_x x'}\,dx' = \frac{2a}{\pi}\,\frac{\cos(k_x a/2)}{1-(k_x a/\pi)^2}$$

It follows that:

$$\boxed{f_y(\theta,\phi) = E_0\,\frac{2ab}{\pi}\,\frac{\cos(\pi v_x)}{1-4v_x^2}\,\frac{\sin(\pi v_y)}{\pi v_y}} \tag{18.1.3}$$

where $v_x = k_x a/2\pi$ and $v_y = k_y b/2\pi$, or,

$$\boxed{v_x = \frac{a}{\lambda}\sin\theta\cos\phi\,,\quad v_y = \frac{b}{\lambda}\sin\theta\sin\phi} \tag{18.1.4}$$

The obliquity factors can be chosen to be one of the three cases: (a) the PEC case, if the aperture is terminated in a ground plane, (b) the ordinary Huygens source case, if it is radiating into free space, or (c) the modified Huygens source case. Thus,

$$\begin{bmatrix} c_\theta \\ c_\phi \end{bmatrix} = \begin{bmatrix} 1 \\ \cos\theta \end{bmatrix},\quad \frac{1}{2}\begin{bmatrix} 1+\cos\theta \\ 1+\cos\theta \end{bmatrix},\quad \frac{1}{2}\begin{bmatrix} 1+K\cos\theta \\ K+\cos\theta \end{bmatrix} \tag{18.1.5}$$

By normalizing all three cases to unity at $\theta = 0^{\circ}$, we may combine them into:

$$c_E(\theta) = \frac{1+K\cos\theta}{1+K}\,,\quad c_H(\theta) = \frac{K+\cos\theta}{1+K} \tag{18.1.6}$$

where K is one of the three possible values:

$$K = 0\,,\qquad K = 1\,,\qquad K = \frac{\eta}{\eta_{TE}} = \sqrt{1-\left(\frac{\lambda}{2a}\right)^2} \tag{18.1.7}$$

The *normalized gains* along the two principal planes are given as follows. For the xz- or the H-plane, we set $\phi = 0^{\circ}$, which gives $E_\theta = 0$:

$$\boxed{g_H(\theta) = \frac{|E_\phi(\theta)|^2}{|E_\phi|^2_{\max}} = |c_H(\theta)|^2\left|\frac{\cos(\pi v_x)}{1-4v_x^2}\right|^2,\quad v_x = \frac{a}{\lambda}\sin\theta} \tag{18.1.8}$$

And, for the yz- or E-plane, we set $\phi = 90^{\circ}$, which gives $E_\phi = 0$:

$$\boxed{g_E(\theta) = \frac{|E_\theta(\theta)|^2}{|E_\theta|^2_{\max}} = |c_E(\theta)|^2\left|\frac{\sin(\pi v_y)}{\pi v_y}\right|^2,\quad v_y = \frac{b}{\lambda}\sin\theta} \tag{18.1.9}$$

The function $\cos(\pi v_x)/(1-4v_x^2)$ determines the essential properties of the H-plane pattern. It is essentially a double-sinc function, as can be seen from the identity:

$$\frac{\cos(\pi v_x)}{1-4v_x^2} = \frac{\pi}{4}\left[\frac{\sin\left(\pi(v_x+\frac{1}{2})\right)}{\pi(v_x+\frac{1}{2})} + \frac{\sin\left(\pi(v_x-\frac{1}{2})\right)}{\pi(v_x-\frac{1}{2})}\right] \tag{18.1.10}$$

It can be evaluated with the help of the MATLAB function `dsinc`, with usage:

`y = dsinc(x);` % the double-sinc function $\dfrac{\cos(\pi x)}{1-4x^2} = \dfrac{\pi}{4}\left[\mathrm{sinc}(x+0.5)+\mathrm{sinc}(x-0.5)\right]$

The 3-dB width of the E-plane pattern is the same as for the uniform rectangular aperture, $\Delta\theta_y = 0.886\,\lambda/b$. The `dsinc` function has the value $\pi/4$ at $v_x = 1/2$. Its 3-dB point is at $v_x = 0.5945$, its first null at $v_x = 1.5$, and its first sidelobe at $v_x = 1.8894$ and has height 0.0708 or 23 dB down from the main lobe. It follows from $v_x = a\sin\theta/\lambda$ that the 3-dB width in angle space will be $\Delta\theta_x = 2{\times}0.5945\,\lambda/a = 1.189\,\lambda/a$. Thus, the 3-dB widths are in radians and in degrees:

$$\Delta\theta_x = 1.189\,\frac{\lambda}{a} = 68.12^{\circ}\,\frac{\lambda}{a}\,,\quad \Delta\theta_y = 0.886\,\frac{\lambda}{b} = 50.76^{\circ}\,\frac{\lambda}{b} \tag{18.1.11}$$

Example 18.1.1: Fig. 18.1.2 shows the H- and E-plane patterns for a WR90 waveguide operating at 10 GHz, so that $\lambda = 3$ cm. The guide dimensions are $a = 2.282$ cm, $b = 1.016$ cm. The typical MATLAB code for generating these graphs was:

```
a = 2.282; b = 1.016; la = 3;

th = (0:0.5:90) * pi/180;

vx = a/la * sin(th);
vy = b/la * sin(th);

K = sqrt(1 - (la/(2*a))^2);            % alternatively, K = 0, or, K = 1

cE = (1 + K*cos(th))/(K+1);            % normalized obliquity factors
cH = (K + cos(th))/(K+1);

gH = abs(cH .* dsinc(vx).^2);           % uses dsinc
gE = abs(cE .* sinc(vy)).^2;           % uses sinc from SP toolbox

figure; dbp(th,gH,45,12);              dB gain polar plot
figure; dbp(th,gE,45,12);
```

The three choices of obliquity factors have been plotted for comparison. We note that the Huygens source cases, $K = 1$ and $K = \eta/\eta_{TE}$, differ very slightly. The H-plane pattern vanishes at $\theta = 90^{\circ}$ in the PEC case ($K = 0$), but not in the Huygens source cases.

The gain computed from Eq. (18.1.13) is $G = 2.62$ or 4.19 dB, and computed from Eq. (18.1.14), $G = 2.67$ or 4.28 dB, where $K = \eta/\eta_{TE} = 0.75$ and $(K+1)^2/4K = 1.02$.

This waveguide is not a high-gain antenna. Increasing the dimensions a, b is impractical and also would allow the propagation of higher modes, making it very difficult to restrict operation to the TE_{10} mode. □

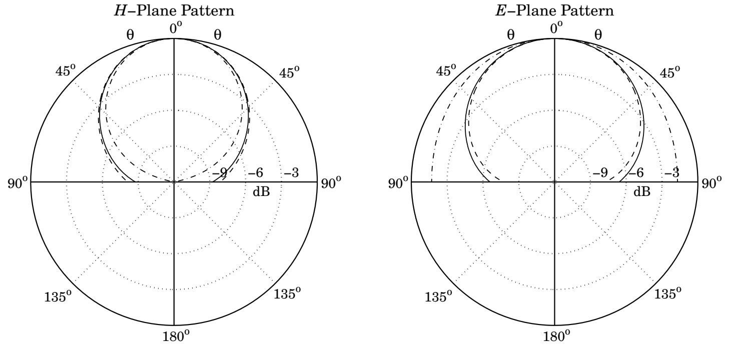


Fig. 18.1.2 Solid line has $K = \eta/\eta_{TE}$, dashed, $K = 1$, and dash-dotted, $K = 0$.

Next, we derive an expression for the directivity and gain of the waveguide aperture. The maximum intensity is obtained at $\theta = 0^{\rm o}$. Because $c_\theta(0) = c_\phi(0)$, we have:

$$U_{\rm max} = \frac{1}{2\eta}\,|\mathbf{E}(0,\phi)|^2 = \frac{1}{2\lambda^2\eta}c_\theta^2(0)\,|f_y(0,\phi)|^2 = \frac{1}{2\lambda^2\eta}c_\theta^2(0)\,|E_0|^2\,\frac{4(ab)^2}{\pi^2}$$

The total power transmitted through the aperture and radiated away is the power propagated down the waveguide given by Eq. (9.7.4), that is,

$$P_{\rm rad} = \frac{1}{4\eta_{TE}}|E_0|^2 ab \tag{18.1.12}$$

It follows that the gain/directivity of the aperture will be:

$$G = 4\pi\,\frac{U_{\rm max}}{P_{\rm rad}} = \frac{4\pi}{\lambda^2}\,\frac{8}{\pi^2}\,(ab)\,\frac{\eta_{TE}}{\eta}\,c_\theta^2(0)$$

For the PEC and ordinary Huygens cases, $c_\theta(0) = 1$. Assuming $\eta_{TE} \simeq \eta$, we have:

$$\boxed{G = \frac{4\pi}{\lambda^2}\,\frac{8}{\pi^2}\,(ab) = 0.81\,\frac{4\pi}{\lambda^2}\,(ab)} \tag{18.1.13}$$

Thus, the effective area of the waveguide aperture is $A_{\rm eff} = 0.81\,(ab)$ and the aperture efficiency $e = 0.81$. For the modified Huygens case, we have for the obliquity factor $c_\theta(0) = (K+1)/2$ with $K = \eta/\eta_{TE}$. It follows that [1162]:

$$G = \frac{4\pi}{\lambda^2}\,\frac{8}{\pi^2}\,(ab)\,\frac{(K+1)^2}{4K} \tag{18.1.14}$$

For waveguides larger than about a wavelength, the directivity factor $(K+1)^2/4K$ is practically equal to unity, and the directivity is accurately given by Eq. (18.1.13). The table below shows some typical values of K and the directivity factor (operation in the TE_{10} mode requires $0.5\lambda < a < \lambda$):

a/λ	K	$(K+1)^2/(4K)$
0.6	0.5528	1.0905
0.8	0.7806	1.0154
1.0	0.8660	1.0052
1.5	0.9428	1.0009
2.0	0.9682	1.0003

The gain-beamwidth product is from Eqs. (18.1.11) and (18.1.13), $p = G\,\Delta\theta_x\,\Delta\theta_y = 4\pi(0.81)\,(0.886)\,(1.189) = 10.723\ \text{rad}^2 = 35\,202\ \text{deg}^2$. Thus, another instance of the general formula (15.3.14) is (with the angles given in radians and in degrees):

$$G = \frac{10.723}{\Delta\theta_x\,\Delta\theta_y} = \frac{35\,202}{\Delta\theta_x^{\rm o}\,\Delta\theta_y^{\rm o}} \tag{18.1.15}$$

18.2 Horn Antennas

The only practical way to increase the directivity of a waveguide is to flare out its ends into a horn. Fig. 18.2.1 shows three types of horns: The H-plane sectoral horn in which the long side of the waveguide (the a-side) is flared, the E-plane sectoral horn in which the short side is flared, and the pyramidal horn in which both sides are flared.

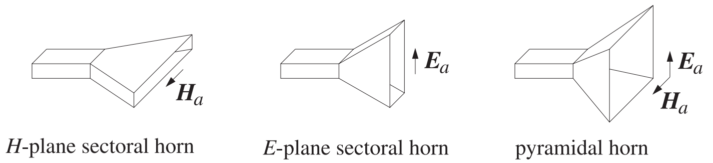


Fig. 18.2.1 H-plane, E-plane, and pyramidal horns.

The pyramidal horn is the most widely used antenna for feeding large microwave dish antennas and for calibrating them. The sectoral horns may be considered as special limits of the pyramidal horn. We will discuss only the pyramidal case.

Fig. 18.2.2 shows the geometry in more detail. The two lower figures are the cross-sectional views along the xz- and yz-planes. It follows from the geometry that the various lengths and flare angles are given by:

$$\begin{aligned}
R_a &= \frac{A}{A-a}R_A\,, & R_b &= \frac{B}{B-b}R_B \\
L_a^2 &= R_a^2 + \frac{A^2}{4}\,, & L_b^2 &= R_b^2 + \frac{B^2}{4} \\
\tan\alpha &= \frac{A}{2R_a}\,, & \tan\beta &= \frac{B}{2R_b} \\
\Delta_a &= \frac{A^2}{8R_a}\,, & \Delta_b &= \frac{B^2}{8R_b}
\end{aligned} \tag{18.2.1}$$

The quantities R_A and R_B represent the perpendicular distances from the plane of the waveguide opening to the plane of the horn. Therefore, they must be *equal*, $R_A = R_B$. Given the horn sides A, B and the common length R_A, Eqs. (18.2.1) allow the calculation of all the relevant geometrical quantities required for the construction of the horn.

The lengths Δ_a and Δ_b represent the maximum deviation of the radial distance from the plane of the horn. The expressions given in Eq. (18.2.1) are approximations obtained when $R_a \gg A$ and $R_b \gg B$. Indeed, using the small-x expansion,

$$\sqrt{1\pm x} \simeq 1 \pm \frac{x}{2}$$

we have two possible ways to approximate Δ_a:

$$\begin{aligned}\Delta_a = L_a - R_a &= \sqrt{R_a^2+\frac{A^2}{4}} - R_a = R_a\sqrt{1+\frac{A^2}{4R_a^2}} - R_a \simeq \frac{A^2}{8R_a}\\ &= L_a - \sqrt{L_a^2-\frac{A^2}{4}} = L_a - L_a\sqrt{1-\frac{A^2}{4L_a^2}} \simeq \frac{A^2}{8L_a}\end{aligned} \tag{18.2.2}$$

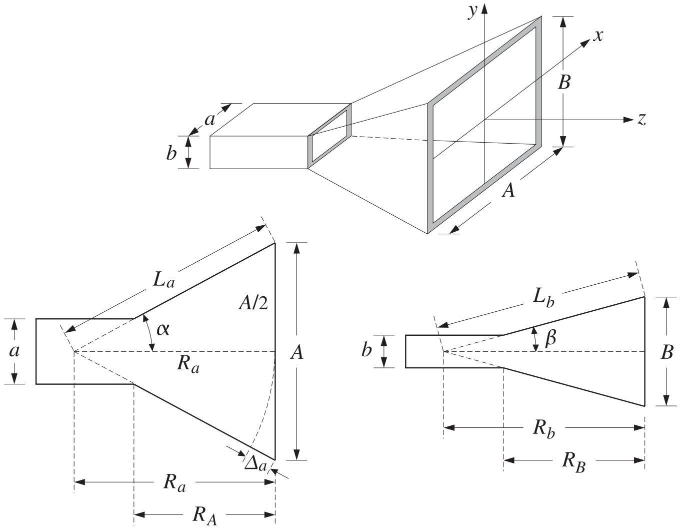


Fig. 18.2.2 The geometry of the pyramidal horn requires $R_A = R_B$.

The two expressions are equal to within the assumed approximation order. The length Δ_a is the maximum deviation of the radial distance at the edge of the horn plane, that is, at $x = \pm A/2$. For any other distance x along the A-side of the horn, and distance y along the B-side, the deviations will be:

$$\Delta_a(x) = \frac{x^2}{2R_a}\,, \quad \Delta_b(y) = \frac{y^2}{2R_b} \tag{18.2.3}$$

The quantities $k\Delta_a(x)$ and $k\Delta_b(y)$ are the relative phase differences at the point (x,y) on the aperture of the horn relative to the center of the aperture. To account for these phase differences, the aperture electric field is assumed to have the form:

$$E_y(x,y) = E_0\cos\left(\frac{\pi x}{A}\right)e^{-jk\Delta_a(x)}\,e^{-jk\Delta_b(y)}\,, \quad \text{or,} \tag{18.2.4}$$

$$\boxed{E_y(x,y) = E_0\cos\left(\frac{\pi x}{A}\right)e^{-jkx^2/2R_a}\,e^{-jky^2/2R_b}} \tag{18.2.5}$$

We note that at the connecting end of the waveguide the electric field is $E_y(x,y) = E_0\cos(\pi x/a)$ and changes gradually into the form of Eq. (18.2.5) at the horn end.

Because the aperture sides A, B are assumed to be large compared to λ, the Huygens source assumption is fairly accurate for the tangential aperture magnetic field, $H_x(x,y) = -E_y(x,y)/\eta$, so that:

$$H_x(x,y) = -\frac{1}{\eta}\,E_0\cos\left(\frac{\pi x}{A}\right)e^{-jkx^2/2R_a}\,e^{-jky^2/2R_b} \tag{18.2.6}$$

The quantities $k\Delta_a$, $k\Delta_b$ are the maximum phase deviations in radians. Therefore, Δ_a/λ and Δ_b/λ will be the maximum deviations in cycles. We define:

$$S_a = \frac{\Delta_a}{\lambda} = \frac{A^2}{8\lambda R_a}\,, \quad S_b = \frac{\Delta_b}{\lambda} = \frac{B^2}{8\lambda R_b} \tag{18.2.7}$$

It turns out that the optimum values of these parameters that result into the highest directivity are approximately: $S_a = 3/8$ and $S_b = 1/4$. We will use these values later in the design of optimum horns. For the purpose of deriving convenient expressions for the radiation patterns of the horn, we define the related quantities:

$$\boxed{\sigma_a^2 = 4S_a = \frac{A^2}{2\lambda R_a}\,, \quad \sigma_b^2 = 4S_b = \frac{B^2}{2\lambda R_b}} \tag{18.2.8}$$

The near-optimum values of these constants are $\sigma_a = \sqrt{4S_a} = \sqrt{4(3/8)} = 1.2247$ and $\sigma_b = \sqrt{4S_b} = \sqrt{4(1/4)} = 1$. These are used very widely, but they are not quite the true optimum values, which are $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$.

Replacing $k = 2\pi/\lambda$ and $2\lambda R_a = A^2/\sigma_a^2$ and $2\lambda R_b = B^2/\sigma_b^2$ in Eq. (18.2.5), we may rewrite the aperture fields in the form: For $-A/2 \le x \le A/2$ and $-B/2 \le y \le B/2$,

$$\boxed{\begin{aligned}E_y(x,y) &= E_0\cos\left(\frac{\pi x}{A}\right)e^{-j(\pi/2)\sigma_a^2(2x/A)^2}\,e^{-j(\pi/2)\sigma_b^2(2y/B)^2}\\ H_x(x,y) &= -\frac{1}{\eta}\,E_0\cos\left(\frac{\pi x}{A}\right)e^{-j(\pi/2)\sigma_a^2(2x/A)^2}\,e^{-j(\pi/2)\sigma_b^2(2y/B)^2}\end{aligned}} \tag{18.2.9}$$

18.3 Horn Radiation Fields

As in the case of the open-ended waveguide, the aperture Fourier transform of the electric field has only a y-component given by:

$$f_y(\theta,\phi)=\int_{-A/2}^{A/2}\int_{-B/2}^{B/2}E_y(x,y)\,e^{jk_xx+jk_yy}\,dx\,dy$$

$$=E_0\int_{-A/2}^{A/2}\cos\left(\frac{\pi x}{A}\right)e^{jk_xx}e^{-j(\pi/2)\sigma_a^2(2x/A)^2}\,dx\cdot\int_{-B/2}^{B/2}e^{jk_yy}e^{-j(\pi/2)\sigma_b^2(2y/B)^2}\,dy$$

The above integrals can be expressed in terms of the following diffraction-like integrals, whose properties are discussed in Appendix F:

$$\boxed{\begin{aligned}F_0(\nu,\sigma)&=\int_{-1}^{1}e^{j\pi\nu\xi}\,e^{-j(\pi/2)\sigma^2\xi^2}\,d\xi\\ F_1(\nu,\sigma)&=\int_{-1}^{1}\cos\left(\frac{\pi\xi}{2}\right)e^{j\pi\nu\xi}\,e^{-j(\pi/2)\sigma^2\xi^2}\,d\xi\end{aligned}}\tag{18.3.1}$$

The function $F_0(\nu,\sigma)$ can be expressed as:

$$\boxed{F_0(\nu,\sigma)=\frac{1}{\sigma}\,e^{j(\pi/2)\,(\nu^2/\sigma^2)}\,\left[F\left(\frac{\nu}{\sigma}+\sigma\right)-F\left(\frac{\nu}{\sigma}-\sigma\right)\right]}\tag{18.3.2}$$

where $F(x)=C(x)-jS(x)$ is the standard Fresnel integral, discussed in Appendix F. Then, the function $F_1(\nu,\sigma)$ can be expressed in terms of $F_0(\nu,\sigma)$:

$$\boxed{F_1(\nu,\sigma)=\frac{1}{2}\big[F_0(\nu+0.5,\sigma)+F_0(\nu-0.5,\sigma)\big]}\tag{18.3.3}$$

The functions $F_0(\nu,\sigma)$ and $F_1(\nu,s)$ can be evaluated numerically for any vector of values ν and any positive scalar σ (including $\sigma=0$) using the MATLAB function `diffint`, which is further discussed in Appendix F and has usage:

```
F0 = diffint(v,sigma,0);       % evaluates the function F0(ν,σ)
F1 = diffint(v,sigma,1);       % evaluates the function F1(ν,σ)
```

In addition to `diffint`, the following MATLAB functions, to be discussed later, facilitate working with horn antennas:

`hband`	% calculate 3-dB bandedges
`heff`	% calculate aperture efficiency
`hgain`	% calculate *H*- and *E*-plane gains
`hopt`	% optimum horn design
`hsigma`	% calculate optimum values of σ_a,σ_b

Next, we express the radiation patterns in terms of the functions (18.3.1). Defining the normalized wavenumbers $\nu_x=k_xA/2\pi$ and $\nu_y=k_yB/2\pi$, we have:

$$\boxed{\nu_x=\frac{A}{\lambda}\sin\theta\cos\phi\,,\quad \nu_y=\frac{B}{\lambda}\sin\theta\sin\phi}\tag{18.3.4}$$

Changing variables to $\xi=2y/B$, the y-integral can written in terms of $F_0(\nu,\sigma)$:

$$\int_{-B/2}^{B/2}e^{jk_yy}e^{-j(\pi/2)\sigma_b^2(2y/B)^2}\,dy=\frac{B}{2}\int_{-1}^{1}e^{j\pi\nu_y\xi}\,e^{-j(\pi/2)\sigma_b^2\xi^2}\,d\xi=\frac{B}{2}F_0(\nu_y,\sigma_b)$$

Similarly, changing variables to $\xi=2x/A$, we find for the x-integral:

$$\int_{-A/2}^{A/2}\cos\left(\frac{\pi x}{A}\right)e^{jk_xx}e^{-j(\pi/2)\sigma_a^2(2x/A)^2}\,dx$$

$$=\frac{A}{2}\int_{-1}^{1}\cos\left(\frac{\pi\xi}{2}\right)e^{j\pi\nu\xi}\,e^{-j(\pi/2)\sigma_a^2\xi^2}\,d\xi=\frac{A}{2}F_1(\nu_x,\sigma_a)$$

It follows that the Fourier transform $f_y(\theta,\phi)$ will be:

$$\boxed{f_y(\theta,\phi)=E_0\frac{AB}{4}F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)}\tag{18.3.5}$$

The open-ended waveguide and the sectoral horns can be thought of as limiting cases of Eq. (18.3.5), as follows:

1. open-ended waveguide: $\sigma_a=0,\quad A=a,\quad \sigma_b=0,\quad B=b.$
2. *H*-plane sectoral horn: $\sigma_a>0,\quad A>a,\quad \sigma_b=0,\quad B=b.$
3. *E*-plane sectoral horn: $\sigma_a=0,\quad A=a,\quad \sigma_b>0,\quad B>b.$

In these cases, the *F*-factors with $\sigma=0$ can be replaced by the following simplified forms, which follow from equations (F.12) and (F.17) of Appendix F:

$$F_0(\nu_y,0)=2\,\frac{\sin(\pi\nu_y)}{\pi\nu_y}\,,\quad F_1(\nu_x,0)=\frac{4}{\pi}\,\frac{\cos(\pi\nu_x)}{1-4\nu_x^2}\tag{18.3.6}$$

The radiation fields are obtained from Eq. (17.5.5), with obliquity factors $c_\theta(\theta)=c_\phi(\theta)=(1+\cos\theta)/2$. Replacing $k=2\pi/\lambda$, we have:

$$\begin{aligned}E_\theta&=j\,\frac{e^{-jkr}}{\lambda r}\,c_\theta(\theta)\,f_y(\theta,\phi)\sin\phi\\ E_\phi&=j\,\frac{e^{-jkr}}{\lambda r}\,c_\phi(\theta)\,f_y(\theta,\phi)\cos\phi\end{aligned}\tag{18.3.7}$$

or, explicitly,

$$\begin{aligned}E_\theta&=j\,\frac{e^{-jkr}}{\lambda r}\,E_0\,\frac{AB}{4}\left(\frac{1+\cos\theta}{2}\right)\sin\phi\,F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)\\ E_\phi&=j\,\frac{e^{-jkr}}{\lambda r}\,E_0\,\frac{AB}{4}\left(\frac{1+\cos\theta}{2}\right)\cos\phi\,F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)\end{aligned}\tag{18.3.8}$$

Horn Radiation Patterns

The radiation intensity is $U(\theta,\phi)=r^2\big(|E_\theta|^2+|E_\phi|^2\big)/2\eta$, so that:

$$U(\theta,\phi)=\frac{1}{32\eta\lambda^2}\,|E_0|^2(AB)^2\,c_\theta^2(\theta)\,\big|F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)\big|^2\tag{18.3.9}$$

Assuming that the maximum intensity is towards the forward direction, that is, at $\nu_x = \nu_y = 0$, we have:

$$U_{\max} = \frac{1}{32\eta\lambda^2}\,|E_0|^2(AB)^2\,\big|F_1(0,\sigma_a)\,F_0(0,\sigma_b)\big|^2 \tag{18.3.10}$$

The direction of maximum gain is not necessarily in the forward direction, but it may be nearby. This happens typically when $\sigma_b > 1.54$. Most designs use the optimum value $\sigma_b = 1$, which does have a maximum in the forward direction. With these caveats in mind, we define the *normalized gain*:

$$g(\theta,\phi) = \frac{U(\theta,\phi)}{U_{\max}} = \left|\frac{1+\cos\theta}{2}\right|^2 \left|\frac{F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)}{F_1(0,\sigma_a)\,F_0(0,\sigma_b)}\right|^2 \tag{18.3.11}$$

Similarly, the *H*- and *E*-plane gains corresponding to $\phi = 0^{\text{o}}$ and $\phi = 90^{\text{o}}$ are:

$$\begin{aligned} g_H(\theta) &= \left|\frac{1+\cos\theta}{2}\right|^2 \left|\frac{F_1(\nu_x,\sigma_a)}{F_1(0,\sigma_a)}\right|^2 = g(\theta,0^{\text{o}})\,, \qquad \nu_x = \frac{A}{\lambda}\sin\theta \\ g_E(\theta) &= \left|\frac{1+\cos\theta}{2}\right|^2 \left|\frac{F_0(\nu_y,\sigma_b)}{F_0(0,\sigma_b)}\right|^2 = g(\theta,90^{\text{o}})\,, \qquad \nu_y = \frac{B}{\lambda}\sin\theta \end{aligned} \tag{18.3.12}$$

The normalizing values $F_1(0,\sigma_a)$ and $F_0(0,\sigma_b)$ are obtained from Eqs. (F.11) and (F.15) of Appendix F. They are given in terms of the Fresnel function $F(x) = C(x) - jS(x)$ as follows:

$$\begin{aligned} |F_1(0,\sigma_a)|^2 &= \frac{1}{\sigma_a^2}\left|F\left(\frac{1}{2\sigma_a}+\sigma_a\right) - F\left(\frac{1}{2\sigma_a}-\sigma_a\right)\right|^2 \\ |F_0(0,\sigma_b)|^2 &= 4\left|\frac{F(\sigma_b)}{\sigma_b}\right|^2 \end{aligned} \tag{18.3.13}$$

These have the limiting values for $\sigma_a = 0$ and $\sigma_b = 0$:

$$|F_1(0,0)|^2 = \frac{16}{\pi^2}\,, \quad |F_0(0,0)|^2 = 4 \tag{18.3.14}$$

The mainlobe/sidelobe characteristics of the gain functions $g_H(\theta)$ and $g_E(\theta)$ depend essentially on the two functions:

$$f_1(\nu_x,\sigma_a) = \left|\frac{F_1(\nu_x,\sigma_a)}{F_1(0,\sigma_a)}\right|\,, \quad f_0(\nu_y,\sigma_a) = \left|\frac{F_0(\nu_y,\sigma_b)}{F_0(0,\sigma_b)}\right| \tag{18.3.15}$$

Fig. 18.3.1 shows these functions for the following values of the σ-parameters: $\sigma_a = [0,\ 1.2593,\ 1.37,\ 1.4749,\ 1.54]$ and $\sigma_b = [0,\ 0.7375,\ 1.0246,\ 1.37,\ 1.54]$.

The values $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$ are the optimum values that maximize the horn directivity (they are close to the commonly used values of $\sigma_a = \sqrt{1.5} = 1.2247$ and $\sigma_b = 1$.)

The values $\sigma_a = 1.4749$ and $\sigma_b = \sigma_a/2 = 0.7375$ are the optimum values that achieve the highest directivity for a waveguide and horn that have the same aspect ratio of $b/a = B/A = 1/2$.

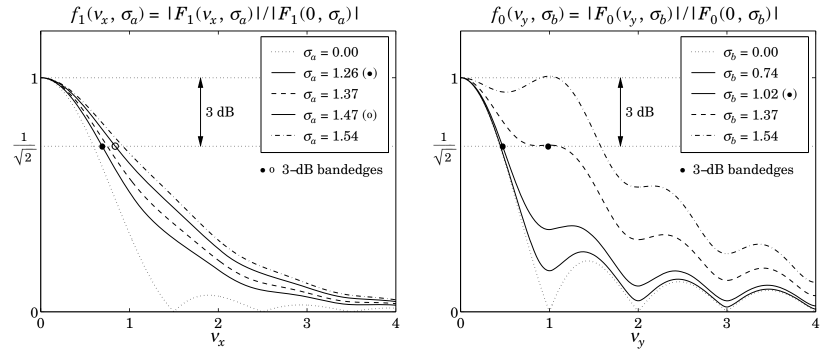


Fig. 18.3.1 Gain functions for different σ-parameters.

For $\sigma_a = \sigma_b = 0$, the functions reduce to the sinc and double-sinc functions of Eq. (18.3.6). The value $\sigma_b = 1.37$ was chosen because the function $f_0(\nu_y,\sigma_b)$ develops a plateau at the 3-dB level, making the definition of the 3-dB width ambiguous.

The value $\sigma_b = 1.54$ was chosen because $f_0(\nu_y,\sigma_b)$ exhibits a secondary maximum away from $\nu_y = 0$. This maximum becomes stronger as σ_b is increased further.

The functions $f_1(\nu,\sigma)$ and $f_0(\nu,\sigma)$ can be evaluated for any vector of ν-values and any σ with the help of the function `diffint`. For example, the following code computes them over the interval $0 \le \nu \le 4$ for the optimum values $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$, and also determines the 3-dB bandedges with the help of the function `hband`:

```
sa = 1.2593; sb = 1.0249;
v = 0:0.01:4;

f1 = abs(diffint(v,sa,1) / diffint(0,sa,1));
f0 = abs(diffint(v,sb,0) / diffint(0,sb,0));

va = hband(sa,1);                    % 3-dB bandedge for H-plane pattern
vb = hband(sb,0);                    % 3-dB bandedge for E-plane pattern
```

The mainlobes become wider as σ_a and σ_b increase. The 3-dB bandedges corresponding to the optimum σs are found from `hband` to be $\nu_a = 0.6928$ and $\nu_b = 0.4737$, and are shown on the graphs.

The 3-dB width in angle θ can be determined from $\nu_x = (A/\lambda)\sin\theta$, which gives approximately $\Delta\theta_a = (2\nu_a)(\lambda/A)$—the approximation being good for $A > 2\lambda$. Thus, in radians and in degrees, we obtain the *H*-plane and *E*-plane optimum 3-dB widths:

$$\Delta\theta_a = 1.3856\,\frac{\lambda}{A} = 79.39^{\text{o}}\,\frac{\lambda}{A}\,, \quad \Delta\theta_b = 0.9474\,\frac{\lambda}{B} = 54.28^{\text{o}}\,\frac{\lambda}{B} \tag{18.3.16}$$

The indicated angles must be replaced by 77.90° and 53.88° if the near-optimum σs are used instead, that is, $\sigma_a = 1.2247$ and $\sigma_b = 1$.

Because of the 3-dB plateau of $f_0(\nu_y,\sigma_b)$ at or near $\sigma_b = 1.37$, the function `hband` defines the bandedge to be in the middle of the plateau. At $\sigma_b = 1.37$, the computed bandedge is $\nu_b = 0.9860$, and is shown in Fig. 18.3.1.

The 3-dB bandedges for the parameters $\sigma_a = 1.4749$ and $\sigma_b = 0.7375$ corresponding to aspect ratio of $1/2$ are $v_a = 0.8402$ (shown on the left graph) and $v_b = 0.4499$.

The MATLAB function `hgain` computes the gains $g_H(\theta)$ and $g_E(\theta)$ at $N+1$ equally spaced angles over the interval $[0, \pi/2]$, given the horn dimensions A, B and the parameters σ_a, σ_b. It has usage:

```
[gh,ge,th] = hgain(N,A,B,sa,sb);    % note: th = linspace(0, pi/2, N+1)
[gh,ge,th] = hgain(N,A,B);          % uses optimum values σa = 1.2593, σb = 1.0246
```

Example 18.3.1: Fig. 18.3.2 shows the *H*- and *E*-plane gains of a horn with sides $A = 4\lambda$ and $B = 3\lambda$ and for the optimum values of the σ-parameters. The 3-dB angle widths were computed from Eq. (18.3.16) to be: $\Delta\theta_a = 19.85^{\circ}$ and $\Delta\theta_b = 18.09^{\circ}$.

The graphs show also a 3-dB gain circle as it intersects the gain curves at the 3-dB angles, which are $\Delta\theta_a/2$ and $\Delta\theta_b/2$.

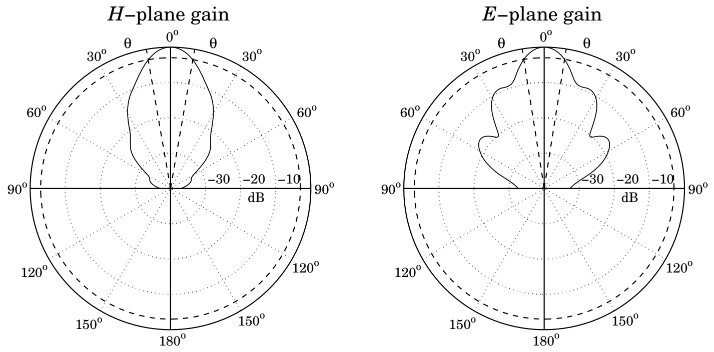


Fig. 18.3.2 *H*- and *E*-plane gains for $A = 4\lambda$, $B = 3\lambda$, and $\sigma_a = 1.2593$, $\sigma_b = 1.0246$.

The essential MATLAB code for generating the left graph was:

```
A = 4; B = 3; N = 200;

[gh,ge,th] = hgain(N,A,B);          % calculate gains

Dtha = 79.39/A;                     % calculate width Δθa

dbp(th,gh);                         % make polar plot in dB
addbwp(Dtha);                       % add the 3-dB widths
addcirc(3);                         % add a 3-dB gain circle
```

We will see later that the gain of this horn is $G = 18.68$ dB and that it can fit on a waveguide with sides $a = \lambda$ and $b = 0.35\lambda$, with an axial length of $R_A = R_B = 3.78\lambda$. □

18.4 Horn Directivity

The radiated power P_{rad} is obtained by integrating the Poynting vector of the aperture fields over the horn area. The quadratic phase factors in Eq. (18.2.9) have no effect on this calculation, the result being the same as in the case of a waveguide. Thus,

$$P_{\text{rad}} = \frac{1}{4\eta}|E_0|^2(AB) \tag{18.4.1}$$

It follows that the horn directivity will be:

$$\boxed{G = 4\pi\frac{U_{\max}}{P_{\text{rad}}} = \frac{4\pi}{\lambda^2}(AB)\frac{1}{8}\big|F_1(0,\sigma_a)\,F_0(0,\sigma_b)\big|^2 = e\,\frac{4\pi}{\lambda^2}AB} \tag{18.4.2}$$

where we defined the aperture efficiency e by:

$$\boxed{e(\sigma_a,\sigma_b) = \frac{1}{8}\big|F_1(0,\sigma_a)\,F_0(0,\sigma_b)\big|^2} \tag{18.4.3}$$

Using the MATLAB function `diffint`, we may compute e for any values of σ_a, σ_b. In particular, we find for the optimum values $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$:

$$\begin{aligned} \sigma_a = 1.2593 \quad &\Rightarrow \quad |F_1(0,\sigma_a)|^2 = |\texttt{diffint}(0,\sigma_a,1)|^2 = 1.2520 \\ \sigma_b = 1.0246 \quad &\Rightarrow \quad |F_0(0,\sigma_b)|^2 = |\texttt{diffint}(0,\sigma_b,0)|^2 = 3.1282 \end{aligned} \tag{18.4.4}$$

This leads to the aperture efficiency:

$$e = \frac{1}{8}(1.2520)(3.1282) \simeq 0.49 \tag{18.4.5}$$

and to the optimum horn directivity:

$$\boxed{G = 0.49\,\frac{4\pi}{\lambda^2}AB} \qquad \text{(optimum horn directivity)} \tag{18.4.6}$$

If we use the near-optimum values of $\sigma_a = \sqrt{1.5}$ and $\sigma_b = 1$, the calculated efficiency becomes $e = 0.51$. It may seem strange that the efficiency is larger for the non-optimum σ_a, σ_b. We will see in the next section that "optimum" does not mean maximizing the efficiency, but rather maximizing the gain given the geometrical constraints of the horn.

The gain-beamwidth product is from Eqs. (18.3.16) and (18.4.6), $p = G\,\Delta\theta_a\,\Delta\theta_b = 4\pi(0.49)(1.3856)(0.9474) = 8.083\ \text{rad}^2 = 26\,535\ \text{deg}^2$. Thus, in radians and in degrees, we have another instance of (15.3.14):

$$G = \frac{8.083}{\Delta\theta_a\,\Delta\theta_b} = \frac{26\,535}{\Delta\theta_a^{o}\,\Delta\theta_b^{o}} \tag{18.4.7}$$

The gain of the *H*-plane sectoral horn is obtained by setting $\sigma_b = 0$, which gives $F_0(0,0) = 2$. Similarly, the *E*-plane horn is obtained by setting $\sigma_a = 0$, with $F_1(0,0) = 4/\pi$. Thus, we have:

$$\begin{aligned} G_H &= \frac{4\pi}{\lambda^2}(AB)\frac{1}{8}|F_1(0,\sigma_a)|^2 4 = \frac{2\pi}{\lambda^2}(AB)\,|F_1(0,\sigma_a)|^2 \\ G_E &= \frac{4\pi}{\lambda^2}(AB)\frac{1}{8}\frac{16}{\pi^2}|F_0(0,\sigma_b)|^2 = \frac{8}{\pi\lambda^2}(AB)\,|F_0(0,\sigma_b)|^2 \end{aligned} \tag{18.4.8}$$

The corresponding aperture efficiencies follow by dividing Eqs. (18.4.8) by $4\pi AB/\lambda^2$:

$$e_H(\sigma_a)=e(\sigma_a,0)=\frac{1}{2}\left|F_1(0,\sigma_a)\right|^2,\quad e_E(\sigma_b)=e(0,\sigma_b)=\frac{2}{\pi^2}\left|F_0(0,\sigma_b)\right|^2$$

In the limit $\sigma_a = \sigma_b = 0$, we find $e = 0.81$, which agrees with Eq. (18.1.13) of the open waveguide case. The MATLAB function `heff` calculates the aperture efficiency $e(\sigma_a,\sigma_b)$ for any values of σ_a, σ_b. It has usage:

```
e = heff(sa,sb);             % horn antenna efficiency
```

Next, we discuss the conditions for optimum directivity. In constructing a horn antenna, we have the constraints of (a) keeping the dimensions a,b of the feeding waveguide small enough so that only the TE_{10} mode is excited, and (b) maintaining the equality of the axial lengths $R_A = R_B$ between the waveguide and horn planes, as shown in Fig. 18.2.2. Using Eqs. (18.2.1) and (18.2.8), we have:

$$R_A = \frac{A-a}{A}R_a = \frac{A(A-a)}{2\lambda\sigma_a^2},\qquad R_B = \frac{B-b}{B}R_b = \frac{B(B-b)}{2\lambda\sigma_b^2} \tag{18.4.9}$$

Then, the geometrical constraint $R_A = R_B$ implies;

$$\frac{A(A-a)}{2\lambda\sigma_a^2} = \frac{B(B-b)}{2\lambda\sigma_b^2}\quad\Rightarrow\quad\boxed{\frac{\sigma_b^2}{\sigma_a^2} = \frac{B(B-b)}{A(A-a)}} \tag{18.4.10}$$

We wish to maximize the gain while respecting the geometry of the horn. For a fixed axial distance $R_A = R_B$, we wish to determine the optimum dimensions A,B that will maximize the gain.

The lengths R_A, R_B are related to the radial lengths R_a, R_b by Eq. (18.4.9). For $A \gg a$, the lengths R_a and R_A are practically equal, and similarly for R_b and R_B. Therefore, an almost equivalent (but more convenient) problem is to find A,B that maximize the gain for fixed values of the radial distances R_a, R_b.

Because of the relationships $A = \sigma_a\sqrt{2\lambda R_a}$ and $B = \sigma_b\sqrt{2\lambda R_b}$, this problem is equivalent to finding the optimum values of σ_a and σ_b that will maximize the gain. Replacing A,B in Eq. (18.4.2), we rewrite G in the form:

$$G = \frac{4\pi}{\lambda^2}\left(\sigma_a\sqrt{2\lambda R_a}\right)\left(\sigma_b\sqrt{2\lambda R_b}\right)\frac{1}{8}\left|F_1(0,\sigma_a)\,F_0(0,\sigma_b)\right|^2,\quad\text{or,}$$

$$G = \frac{\pi\sqrt{R_aR_b}}{\lambda}f_a(\sigma_a)f_b(\sigma_b) \tag{18.4.11}$$

where we defined the directivity functions:

$$f_a(\sigma_a)=\sigma_a\left|F_1(0,\sigma_a)\right|^2,\quad f_b(\sigma_b)=\sigma_b\left|F_0(0,\sigma_b)\right|^2 \tag{18.4.12}$$

These functions are plotted on the left graph of Fig. 18.4.1. Their maxima occur at $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$. As we mentioned before, these values are sometimes approximated by $\sigma_a = \sqrt{1.5} = 1.2244$ and $\sigma_b = 1$.

An alternative class of directivity functions can be derived by constructing a horn whose aperture has the *same* aspect ratio as the waveguide, that is,

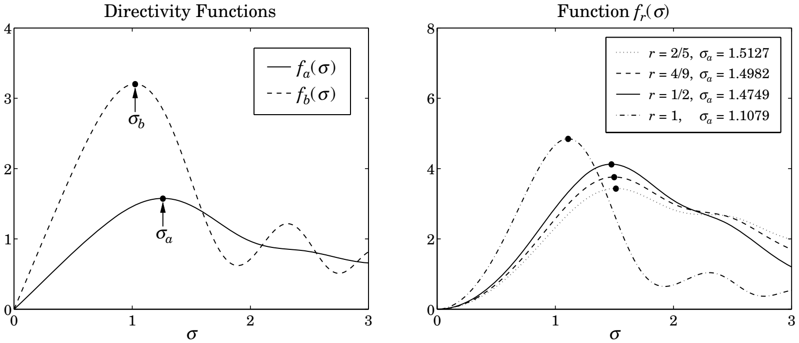


Fig. 18.4.1 Directivity functions.

$$\frac{B}{A} = \frac{b}{a} = r \tag{18.4.13}$$

The aspect ratio of a typical waveguide is of the order of $r = 0.5$, which ensures the largest operating bandwidth in the TE_{10} mode and the largest power transmitted.

It follows from Eq. (18.4.13) that (18.4.10) will be satisfied provided $\sigma_b^2/\sigma_a^2 = r^2$, or $\sigma_b = r\sigma_a$. The directivity (18.4.11) becomes:

$$G = \frac{\pi\sqrt{R_aR_b}}{\lambda}f_r(\sigma_a) \tag{18.4.14}$$

where we defined the function:

$$f_r(\sigma_a)=f_a(\sigma_a)f_b(r\sigma_a)=r\,\sigma_a^2\left|F_1(0,\sigma_a)F_0(0,r\sigma_a)\right|^2 \tag{18.4.15}$$

This function has a maximum, which depends on the aspect ratio r. The right graph of Fig. 18.4.1 shows $f_r(\sigma)$ and its maxima for various values of r. The aspect ratio $r = 1/2$ is used in many standard guides, $r = 4/9$ is used in the WR-90 waveguide, and $r = 2/5$ in the WR-42.

The MATLAB function `hsigma` computes the optimum σ_a and $\sigma_b = r\sigma_a$ for a given aspect ratio r. It has usage:

```
[sa,sb] = hsigma(r);          % optimum σ-parameters
```

With input $r = 0$, it outputs the separate optimal values $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$. For $r = 0.5$, it gives $\sigma_a = 1.4749$ and $\sigma_b = \sigma_a/2 = 0.7375$, with corresponding aperture efficiency $e = 0.4743$.

18.5 Horn Design

The design problem for a horn antenna is to determine the sides A,B that will achieve a given gain G and will also fit geometrically with a given waveguide of sides a,b, satisfying the condition $R_A = R_B$. The two design equations for A, B are then Eqs. (18.4.2) and (18.4.10):

$$\boxed{G = e\,\frac{4\pi}{\lambda^2}\,AB\,,\qquad \frac{\sigma_b^2}{\sigma_a^2} = \frac{B(B-b)}{A(A-a)}} \tag{18.5.1}$$

The design of the constant aspect ratio case is straightforward. Because $\sigma_b = r\sigma_a$, the second condition is already satisfied. Then, the first condition can be solved for A, from which one obtains $B = rA$ and $R_A = A(A-a)/(2\lambda\sigma_a^2)$:

$$G = e\,\frac{4\pi}{\lambda^2}\,A(rA) \quad\Rightarrow\quad \boxed{A = \lambda\sqrt{\frac{G}{4\pi e r}}} \tag{18.5.2}$$

In Eq. (18.5.2), the aperture efficiency e must be calculated from Eq. (18.4.3) with the help of the MATLAB function `heff`.

For unequal aspect ratios and arbitrary σ_a, σ_b, one must solve the system of equations (18.5.1) for the two unknowns A, B. To avoid negative solutions for B, the second equation in (18.5.1) can be solved for B in terms of A, a, b, thus replacing the above system with:

$$\begin{aligned} f_1(A,B) &\equiv B - \left[\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{\sigma_b^2}{\sigma_a^2}\,A(A-a)}\,\right] = 0 \\ f_2(A,B) &\equiv AB - \frac{\lambda^2 G}{4\pi e} = 0 \end{aligned} \tag{18.5.3}$$

This system can be solved iteratively using Newton's method, which amounts to starting with some initial values A, B and keep replacing them with the corrected values $A + \Delta A$ and $B + \Delta B$, where the corrections are computed from:

$$\begin{bmatrix} \Delta A \\ \Delta B \end{bmatrix} = -\mathcal{M}^{-1}\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \quad \text{where} \quad \mathcal{M} = \begin{bmatrix} \partial_A f_1 & \partial_B f_1 \\ \partial_A f_2 & \partial_B f_2 \end{bmatrix}$$

The matrix $\mathcal{M}$ is given by:

$$\mathcal{M} = \begin{bmatrix} -\dfrac{\sigma_b^2}{\sigma_a^2}\,\dfrac{2A-a}{(2B-b-2f_1)} & 1 \\ B & A \end{bmatrix} \simeq \begin{bmatrix} -\dfrac{\sigma_b^2}{\sigma_a^2}\,\dfrac{2A-a}{2B-b} & 1 \\ B & A \end{bmatrix}$$

where we replaced the $2f_1$ term by zero (this is approximately correct near convergence.) Good initial values are obtained by assuming that A, B will be much larger than a, b and therefore, we write Eq. (18.5.1) approximately in the form:

$$G = e\,\frac{4\pi}{\lambda^2}\,AB\,,\qquad \frac{\sigma_b^2}{\sigma_a^2} = \frac{B^2}{A^2} \tag{18.5.4}$$

This system can be solved easily, giving the initial values:

$$A_0 = \lambda\sqrt{\frac{G}{4\pi e}\,\frac{\sigma_a}{\sigma_b}}\,,\quad B_0 = \lambda\sqrt{\frac{G}{4\pi e}\,\frac{\sigma_b}{\sigma_a}} \tag{18.5.5}$$

Note that these are the same solutions as in the constant-r case. The algorithm converges extremely fast, requiring about 3-5 iterations. It has been implemented by the MATLAB function `hopt` with usage:

```
[A,B,R,err] = hopt(G,a,b,sa,sb);          % optimum horn antenna design
[A,B,R,err] = hopt(G,a,b,sa,sb,N);        % N is the maximum number of iterations
[A,B,R,err] = hopt(G,a,b,sa,sb,0);        % outputs initial values only
```

where G is the desired gain in dB, a, b are the waveguide dimensions. The output R is the common axial length $R = R_A = R_B$. All lengths are given in units of λ. If the parameters σ_a, σ_b are omitted, their optimum values are used. The quantity `err` is the approximation error, and N, the maximum number of iterations (default is 10.)

Example 18.5.1: Design a horn antenna with gain 18.68 dB and waveguide sides of $a = \lambda$ and $b = 0.35\lambda$. The following call to `hopt`,

```
[A,B,R,err] = hopt(18.68, 1, 0.35);
```

yields the values (in units of λ): $A = 4$, $B = 2.9987$, $R = 3.7834$, and $err = 3.7\times 10^{-11}$. These are the same as in Example 18.3.1. □

Example 18.5.2: Design a horn antenna operating at 10 GHz and fed by a WR-90 waveguide with sides $a = 2.286$ cm and $b = 1.016$ cm. The required gain is 23 dB ($G = 200$).

Solution: The wavelength is $\lambda = 3$ cm. We carry out two designs, the first one using the optimum values $\sigma_a = 1.2593$, $\sigma_b = 1.0246$, and the second using the aspect ratio of the WR-90 waveguide, which is $r = b/a = 4/9$, and corresponds to $\sigma_a = 1.4982$ and $\sigma_b = 0.6659$. The following MATLAB code calculates the horn sides for the two designs and plots the E-plane patterns:

```
la = 3; a = 2.286; b = 1.016;                         % lengths in cm
G = 200; Gdb = 10*log10(G);                           % G_dB = 23.0103 dB

[sa1,sb1] = hsigma(0);                                % optimum σ-parameters
[A1,B1,R1] = hopt(Gdb, a/la, b/la, sa1, sb1);         % A1, B1, R1 in units of λ

[sa2,sb2] = hsigma(b/a);                              % optimum σ's for r = b/a
[A2,B2,R2] = hopt(Gdb, a/la, b/la, sa2, sb2,0);       % output initial values

N = 200;                                              % 201 angles in 0 ≤ θ ≤ π/2

[gh1,ge1,th] = hgain(N,A1,B1,sa1,sb1);                % calculate gains
[gh2,ge2,th] = hgain(N,A2,B2,sa2,sb2);

figure; dbp(th,gh1); figure; dbp(th,ge1);             % polar plots in dB
figure; dbp(th,gh2); figure; dbp(th,ge2);

A1 = A1*la; B1 = B1*la; R1 = R1*la;                   % lengths in cm
A2 = A2*la; B2 = B2*la; R2 = R2*la;
```

The designed sides and axial lengths are in the two cases:

$$\begin{aligned} &A_1 = 19.2383 \text{ cm}, \quad B_1 = 15.2093 \text{ cm}, \quad R_1 = 34.2740 \text{ cm} \\ &A_2 = 26.1457 \text{ cm}, \quad B_2 = 11.6203 \text{ cm}, \quad R_2 = 46.3215 \text{ cm} \end{aligned}$$

The H- and E-plane patterns are plotted in Fig. 18.5.1. The first design (top graphs) has slightly wider 3-dB width in the H-plane because its A-side is shorter than that of the second design. But, its E-plane 3-dB width is narrower because its B-side is longer.

The initial values given in Eq. (18.5.5) can be used to give an alternative, albeit approximate, solution obtained purely algebraically: Compute A_0, B_0, then revise the value of B_0 by recomputing it from the first of Eq. (18.5.3), so that the geometric constraint $R_A = R_B$ is met, and then recompute the gain, which will be slightly different than the required one.

For example, using the optimum values $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$, we find from (18.5.5): $A_0 = 18.9644$, $B_0 = 15.4289$ cm, and $R_A = 33.2401$ cm. Then, we recalculate B_0 to be $B_0 = 13.9453$ cm, and obtain the new gain $G = 180.77$, or, 22.57 dB. □

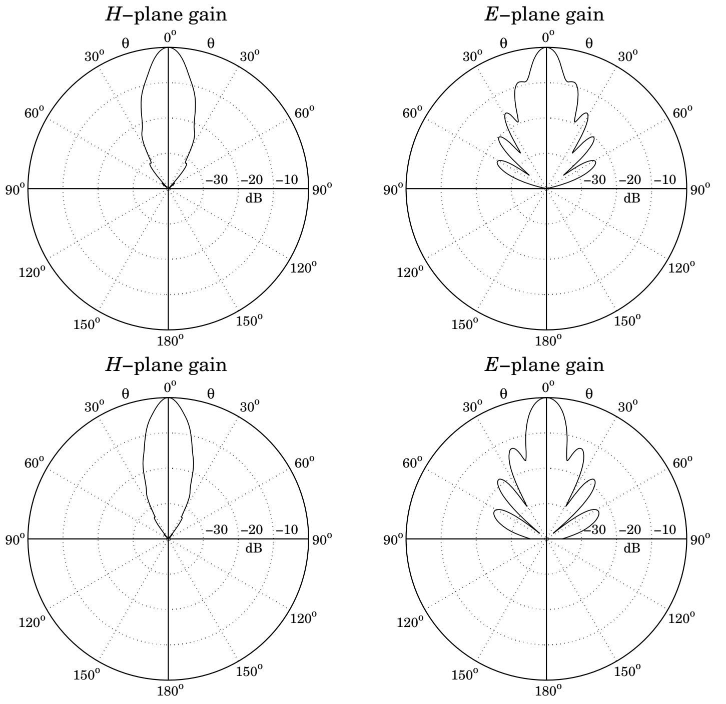


Fig. 18.5.1 *H*- and *E*-plane patterns.

18.6 Microstrip Antennas

A microstrip antenna is a metallic patch on top of a dielectric substrate that sits on top of a ground plane. Fig. 18.6.1 depicts a rectangular microstrip antenna fed by a microstrip line. It can also be fed by a coaxial line, with its inner and outer conductors connected to the patch and ground plane, respectively.

In this section, we consider only rectangular patches and discuss simple aperture models for calculating the radiation patterns of the antenna. Further details and applications of microstrip antennas may be found in [1182–1189].

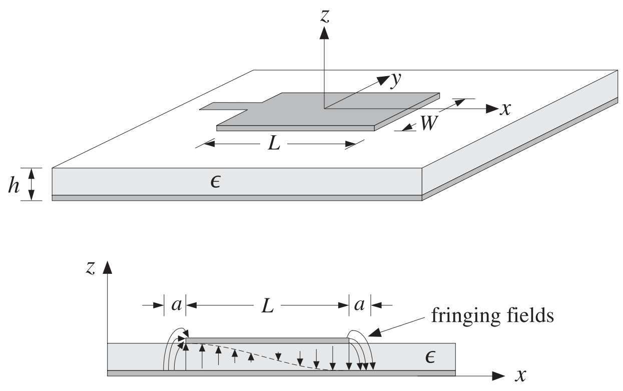


Fig. 18.6.1 Microstrip antenna and *E*-field pattern in substrate.

The height h of the substrate is typically of a fraction of a wavelength, such as $h = 0.05\lambda$, and the length L is of the order of 0.5λ. The structure radiates from the fringing fields that are exposed above the substrate at the edges of the patch.

In the so-called cavity model, the patch acts as resonant cavity with an electric field perpendicular to the patch, that is, along the z-direction. The magnetic field has vanishing tangential components at the four edges of the patch. The fields of the lowest resonant mode (assuming $L \ge W$) are given by:

$$\begin{aligned} E_z(x) &= -E_0 \sin\left(\frac{\pi x}{L}\right) & & -\frac{L}{2} \le x \le \frac{L}{2} \\ & & \text{for} & \\ H_y(x) &= -H_0 \cos\left(\frac{\pi x}{L}\right) & & -\frac{W}{2} \le y \le \frac{W}{2} \end{aligned} \tag{18.6.1}$$

where $H_0 = -jE_0/\eta$. We have placed the origin at the middle of the patch (note that $E_z(x)$ is equivalent to $E_0\cos(\pi x/L)$ for $0 \le x \le L$.)

It can be verified that Eqs. (18.6.1) satisfy Maxwell's equations and the boundary conditions, that is, $H_y(x) = 0$ at $x = \pm L/2$, provided the resonant frequency is:

$$\omega = \frac{\pi c}{L} \quad \Rightarrow \quad f = 0.5\,\frac{c}{L} = 0.5\,\frac{c_0}{L\sqrt{\epsilon_r}} \tag{18.6.2}$$

where $c = c_0/\sqrt{\epsilon_r}$, $\eta = \eta_0/\sqrt{\epsilon_r}$, and ϵ_r is the relative permittivity of the dielectric substrate. It follows that the resonant microstrip length will be half-wavelength:

$$L = 0.5\,\frac{\lambda}{\sqrt{\epsilon_r}} \tag{18.6.3}$$

Fig. 18.6.2 shows two simple models for calculating the radiation patterns of the microstrip antenna. The model on the left assumes that the fringing fields extend over a small distance a around the patch sides and can be replaced with the fields $\boldsymbol{E}_a$ that are tangential to the *substrate surface* [1184]. The four extended edge areas around the patch serve as the effective radiating apertures.

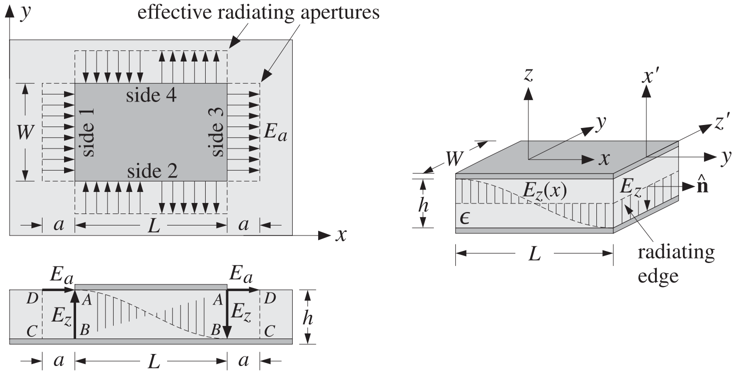


Fig. 18.6.2 Aperture models for microstrip antenna.

The model on the right assumes that the substrate is truncated beyond the extent of the patch [1183]. The four dielectric substrate walls serve now as the radiating apertures. The only tangential aperture field on these walls is $\boldsymbol{E}_a = \hat{\mathbf{z}}\,E_z$, because the tangential magnetic fields vanish by the boundary conditions.

For both models, the ground plane can be eliminated using image theory, resulting in doubling the aperture magnetic currents, that is, $\boldsymbol{J}_{ms} = -2\hat{\mathbf{n}}\times\boldsymbol{E}_a$. The radiation patterns are then determined from $\boldsymbol{J}_{ms}$.

For the first model, the effective tangential fields can be expressed in terms of the field E_z by the relationship: $aE_a = hE_z$. This follows by requiring the vanishing of the line integrals of $\boldsymbol{E}$ around the loops labeled ABCD in the lower left of Fig. 18.6.2. Because $E_z = \pm E_0$ at $x = \pm L/2$, we obtain from the left and right such contours:

$$\oint_{ABCD} \boldsymbol{E}\cdot d\mathbf{l} = -E_0 h + E_a a = 0\,,\qquad \oint_{ABCD} \boldsymbol{E}\cdot d\mathbf{l} = E_0 h - E_a a = 0 \quad\Rightarrow\quad E_a = \frac{hE_0}{a}$$

In obtaining these, we assumed that the electric field is nonzero only along the sides AD and AB. A similar argument for the sides 2 & 4 shows that $E_a = \pm hE_z(x)/a$. The directions of $\boldsymbol{E}_a$ at the four sides are as shown in the figure. Thus, we have:

$$\begin{aligned}
&\text{for sides 1 \& 3}: \quad && \boldsymbol{E}_a = \hat{\mathbf{x}}\,\frac{hE_0}{a}\\
&\text{for sides 2 \& 4}: \quad && \boldsymbol{E}_a = \pm\hat{\mathbf{y}}\,\frac{hE_z(x)}{a} = \mp\hat{\mathbf{y}}\,\frac{hE_0}{a}\,\sin\left(\frac{\pi x}{L}\right)
\end{aligned}\tag{18.6.4}$$

The outward normal to the aperture plane is $\hat{\mathbf{n}} = \hat{\mathbf{z}}$ for all four sides. Therefore, the surface magnetic currents $\boldsymbol{J}_{ms} = -2\hat{\mathbf{n}}\times\boldsymbol{E}_a$ become:

$$\begin{aligned}
&\text{for sides 1 \& 3}: \quad && \boldsymbol{J}_{ms} = \hat{\mathbf{y}}\,\frac{2hE_0}{a}\\
&\text{for sides 2 \& 4}: \quad && \boldsymbol{J}_{ms} = \pm\hat{\mathbf{x}}\,\frac{2hE_0}{a}\,\sin\left(\frac{\pi x}{L}\right)
\end{aligned}\tag{18.6.5}$$

The radiated electric field is obtained from Eq. (17.3.4) by setting $\boldsymbol{F} = 0$ and calculating $\boldsymbol{F}_m$ as the sum of the magnetic radiation vectors over the four effective apertures:

$$\boldsymbol{E} = jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\boldsymbol{F}_m = jk\,\frac{e^{-jkr}}{4\pi r}\,\hat{\mathbf{r}}\times\big[\boldsymbol{F}_{m1}+\boldsymbol{F}_{m2}+\boldsymbol{F}_{m3}+\boldsymbol{F}_{m4}\big]\tag{18.6.6}$$

The vectors $\boldsymbol{F}_m$ are the two-dimensional Fourier transforms over the apertures:

$$\boldsymbol{F}_m(\theta,\phi) = \int_A \boldsymbol{J}_{ms}(x,y)\,e^{jk_x x + jk_y y}\,dS$$

The integration surfaces $dS = dx\,dy$ are approximately, $dS = a\,dy$ for 1 & 3, and $dS = a\,dx$ for 2 & 4. Similarly, in the phase factor $e^{jk_x x + jk_y y}$, we must set $x = \mp L/2$ for sides 1 & 3, and $y = \mp W/2$ for sides 2 & 4. Inserting Eq. (18.6.5) into the Fourier integrals and combining the terms for apertures 1 & 3 and 2 & 4, we obtain:

$$\begin{aligned}
\boldsymbol{F}_{m,13} &= \hat{\mathbf{y}}\,\frac{2hE_0}{a}\int_{-W/2}^{W/2}\big(e^{-jk_xL/2}+e^{jk_xL/2}\big)\,e^{jk_y y}\,a\,dy\\
\boldsymbol{F}_{m,24} &= \hat{\mathbf{x}}\,\frac{2hE_0}{a}\int_{-L/2}^{L/2}\big(e^{-jk_yW/2}-e^{jk_yW/2}\big)\,\sin\left(\frac{\pi x}{L}\right)e^{jk_x x}\,a\,dx
\end{aligned}$$

Note that the a factors cancel. Using Euler's formulas and the integrals:

$$\int_{-W/2}^{W/2} e^{jk_y y}dy = W\,\frac{\sin(k_yW/2)}{k_yW/2}\,,\qquad \int_{-L/2}^{L/2}\sin\left(\frac{\pi x}{L}\right)e^{jk_x x}dx = \frac{2jk_xL^2}{\pi^2}\,\frac{\cos(k_xL/2)}{1-\left(\dfrac{k_xL}{\pi}\right)^2}\,,$$

we find the radiation vectors:

$$\begin{aligned}
\boldsymbol{F}_{m,12} &= \hat{\mathbf{y}}\,4E_0hW\,\cos(\pi v_x)\,\frac{\sin(\pi v_y)}{\pi v_y}\\
\boldsymbol{F}_{m,24} &= \hat{\mathbf{x}}\,4E_0hL\,\frac{4v_x\cos(\pi v_x)}{\pi(1-4v_x^2)}\,\sin(\pi v_y)
\end{aligned}\tag{18.6.7}$$

where we defined the normalized wavenumbers as usual:

$$\boxed{\begin{aligned}
v_x &= \frac{k_xL}{2\pi} = \frac{L}{\lambda}\sin\theta\cos\phi\\
v_y &= \frac{k_yW}{2\pi} = \frac{W}{\lambda}\sin\theta\sin\phi
\end{aligned}}\tag{18.6.8}$$

From Eq. (E.8) of Appendix E, we have:

$$\begin{aligned}
\hat{\mathbf{r}}\times\hat{\mathbf{y}} &= \hat{\mathbf{r}}\times(\hat{\mathbf{r}}\sin\theta\sin\phi + \hat{\boldsymbol{\theta}}\cos\theta\sin\phi + \hat{\boldsymbol{\phi}}\cos\phi) = \hat{\boldsymbol{\phi}}\cos\theta\sin\phi - \hat{\boldsymbol{\theta}}\cos\phi\\
\hat{\mathbf{r}}\times\hat{\mathbf{x}} &= \hat{\mathbf{r}}\times(\hat{\mathbf{r}}\sin\theta\cos\phi + \hat{\boldsymbol{\theta}}\cos\theta\cos\phi - \hat{\boldsymbol{\phi}}\sin\phi) = \hat{\boldsymbol{\phi}}\cos\theta\cos\phi + \hat{\boldsymbol{\theta}}\sin\phi
\end{aligned}$$

It follows from Eq. (18.6.6) that the radiated fields from sides 1 & 3 will be:

$$\boxed{\boldsymbol{E}(\theta,\phi) = jk\,\frac{e^{-jkr}}{4\pi r}\,4E_0hW\,\big[\hat{\boldsymbol{\phi}}\cos\theta\sin\phi - \hat{\boldsymbol{\theta}}\cos\phi\big]\,F(\theta,\phi)}\tag{18.6.9}$$

where we defined the function:

$$\boxed{F(\theta,\phi)=\cos(\pi v_x)\,\frac{\sin(\pi v_y)}{\pi v_y}} \tag{18.6.10}$$

Similarly, we have for sides 2 & 4:

$$\begin{aligned} \mathbf{E}(\theta,\phi) &= jk\,\frac{e^{-jkr}}{4\pi r}\,4E_0hL\,\big[\hat{\boldsymbol{\phi}}\cos\theta\cos\phi+\hat{\boldsymbol{\theta}}\sin\phi\big]\,f(\theta,\phi)\\ f(\theta,\phi) &= \frac{4v_x\cos(\pi v_x)}{\pi(1-4v_x^2)}\,\sin(\pi v_y) \end{aligned} \tag{18.6.11}$$

The normalized gain is found from Eq. (18.6.9) to be:

$$\boxed{g(\theta,\phi)=\frac{|\mathbf{E}(\theta,\phi)|^2}{|\mathbf{E}|^2_{\max}}=\big(\cos^2\theta\sin^2\phi+\cos^2\phi\big)\,\big|F(\theta,\phi)\big|^2} \tag{18.6.12}$$

The corresponding expression for sides 2 & 4, although not normalized, provides a measure for the gain in that case:

$$g(\theta,\phi)=\big(\cos^2\theta\cos^2\phi+\sin^2\phi\big)\,\big|f(\theta,\phi)\big|^2 \tag{18.6.13}$$

The *E*- and *H*-plane gains are obtained by setting $\phi=0^{\mathrm{o}}$ and $\phi=90^{\mathrm{o}}$ in Eq. (18.6.12):

$$\begin{aligned} g_E(\theta) &= \frac{|E_\theta|^2}{|E_\theta|^2_{\max}}=\big|\cos(\pi v_x)\big|^2, && v_x=\frac{L}{\lambda}\sin\theta\\ g_H(\theta) &= \frac{|E_\phi|^2}{|E_\phi|^2_{\max}}=\left|\cos\theta\,\frac{\sin(\pi v_y)}{\pi v_y}\right|^2, && v_y=\frac{W}{\lambda}\sin\theta \end{aligned} \tag{18.6.14}$$

Most of the radiation from the microstrip arises from sides 1 & 3. Indeed, $F(\theta,\phi)$ has a maximum towards broadside, $v_x=v_y=0$, whereas $f(\theta,\phi)$ vanishes. Moreover, $f(\theta,\phi)$ vanishes identically for all θ and $\phi=0^{\mathrm{o}}$ (*E*-plane) or $\phi=90^{\mathrm{o}}$ (*H*-plane).

Therefore, sides 2 & 4 contribute little to the total radiation, and they are usually ignored. For lengths of the order of $L=0.3\lambda$ to $L=\lambda$, the gain function (18.6.13) remains suppressed by 7 to 17 dB for *all directions*, relative to the gain of (18.6.12).

Example 18.6.1: Fig. 18.6.3 shows the *E*- and *H*-plane patterns for $W=L=0.3371\lambda$. Both patterns are fairly broad.

The choice for L comes from the resonant condition $L=0.5\lambda/\sqrt{\epsilon_r}$. For a typical substrate with $\epsilon_r=2.2$, we find $L=0.5\lambda/\sqrt{2.2}=0.3371\lambda$.

Fig. 18.6.4 shows the 3-dimensional gains computed from Eqs. (18.6.12) and (18.6.13). The field strengths (square roots of the gains) are plotted to improve the visibility of the graphs. The MATLAB code for generating these plots was:

```
L = 0.5/sqrt(2.2); W = L;

[th,ph] = meshgrid(0:3:90, 0:6:360); th = th * pi/180; ph = ph * pi/180;

vx = L * sin(th) .* cos(ph);
vy = W * sin(th) .* sin(ph);
```

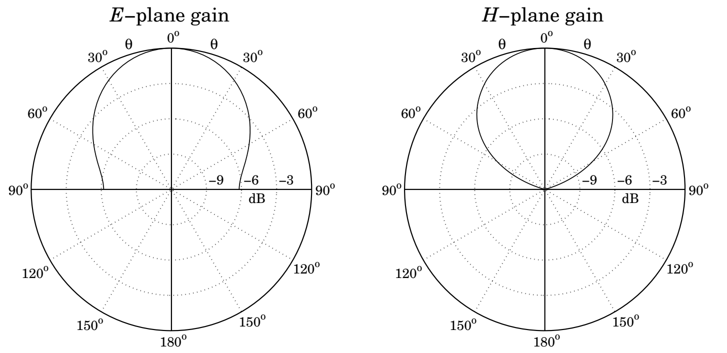


Fig. 18.6.3 *E*- and *H*-plane gains of microstrip antenna.

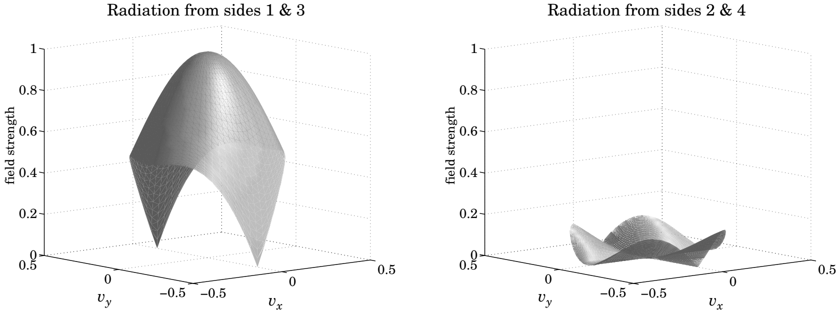


Fig. 18.6.4 Two-dimensional gain patterns from sides 1 & 3 and 2 & 4.

```
E13 = sqrt(cos(th).^2.*sin(ph).^2 + cos(ph).^2);
E13 = E13 .* abs(cos(pi*vx) .* sinc(vy));

figure; surfl(vx,vy,E13);
shading interp; colormap(gray(32));

view([-40,10]);

E24 = sqrt(cos(th).^2.*cos(ph).^2 + sin(ph).^2);
E24 = E24 .* abs(4*vx.*dsinc(vx)/pi .* sin(pi*vy));

figure; surfl(vx,vy,E24);
shading interp; colormap(gray(32));
```

The gain from sides 2 & 4 vanishes along the v_x- and v_y axes, while its maximum in all directions is $\sqrt{g}=0.15$ or -16.5 dB. □

Using the alternative aperture model shown on the right of Fig. 18.6.2, one obtains identical expressions for the magnetic current densities $\mathbf{J}_{ms}$ along the four sides, and therefore, identical radiation patterns. The integration surfaces are now $dS = h\,dy$ for sides 1 & 3, and $dS = h\,dx$ for 2 & 4.

18.7 Parabolic Reflector Antennas

Reflector antennas are characterized by very high gains (30 dB and higher) and narrow main beams. They are widely used in satellite and line-of-sight microwave communications and in radar.

At microwave frequencies, the most common feeds are rectangular, circular, or corrugated horns. Dipole feeds—usually backed by a reflecting plane to enhance their radiation towards the reflector—are used at lower frequencies, typically, up to UHF. Some references on reflector antennas and feed design are [1162–1181].

A typical parabolic reflector, fed by a horn antenna positioned at the focus of the parabola, is shown in Fig. 18.7.1. A geometrical property of parabolas is that all rays originating from the focus get reflected in a direction parallel to the parabola's axis, that is, the z direction.

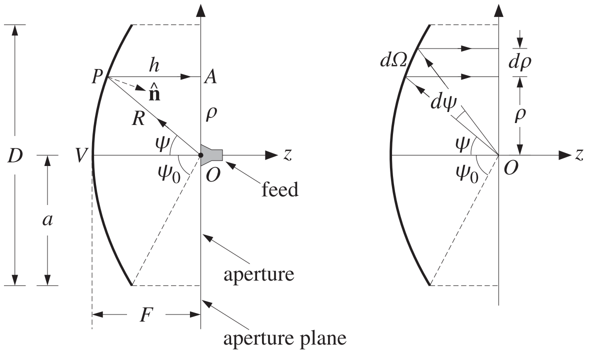


Fig. 18.7.1 Parabolic reflector antenna with feed at the focus.

We choose the origin to be at the focus. An incident ray OP radiated from the feed at an angle ψ becomes the reflected ray PA parallel to the z-axis. The projection of all the reflected rays onto a plane perpendicular to the z-axis—such as the xy-plane—can be considered to be the effective *aperture* of the antenna. This is shown in Fig. 18.7.2.

Let R and h be the lengths of the rays OP and PA. The sum $R + h$ represents the total optical path length from the focus to the aperture plane. This length is constant, independent of ψ, and is given by

$$R + h = 2F \tag{18.7.1}$$

where F is the focal length. The length $2F$ is the total optical length of the incident and reflected axial rays going from O to the vertex V and back to O.

Therefore, all the rays suffer the same phase delay traveling from the focus to the plane. The spherical wave radiated from the feed gets converted upon reflection into a *plane wave*. Conversely, for a receiving antenna, an incident plane wave gets converted into a spherical wave converging onto the focus.

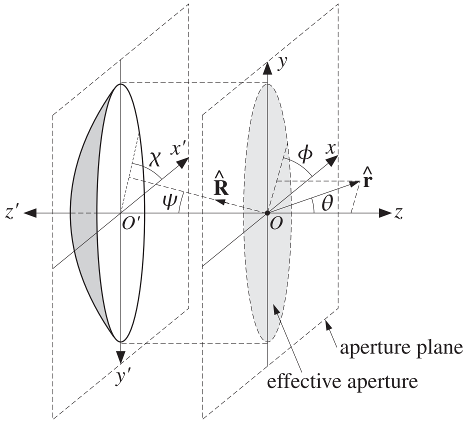


Fig. 18.7.2 Parabolic antenna and its projected effective aperture.

Since $h = R\cos\psi$, Eq. (18.7.1) can be written in the following form, which is the polar representation of the parabolic surface:

$$R + R\cos\psi = 2F \quad\Rightarrow\quad R = \frac{2F}{1+\cos\psi}\,, \qquad \text{or,} \tag{18.7.2}$$

$$\boxed{R = \frac{2F}{1+\cos\psi} = \frac{F}{\cos^2(\psi/2)}} \tag{18.7.3}$$

The radial displacement ρ of the reflected ray on the aperture plane is given by $\rho = R\sin\psi$. Replacing R from (18.7.3), we find:

$$\rho = 2F\,\frac{\sin\psi}{1+\cos\psi} = 2F\tan\left(\frac{\psi}{2}\right) \tag{18.7.4}$$

Similarly, using $R + h = 2F$ or $F - h = R - F$, we have:

$$F - h = F\,\frac{1-\cos\psi}{1+\cos\psi} = F\tan^2\left(\frac{\psi}{2}\right) \tag{18.7.5}$$

It follows that h and ρ will be related by the equation for a parabola:

$$4F(F-h) = \rho^2 \tag{18.7.6}$$

In terms of the xyz-coordinate system, we have $\rho^2 = x^2 + y^2$ and $z = -h$, so that Eq. (18.7.6) becomes the equation for a paraboloid surface:

$$\boxed{4F(F+z) = x^2 + y^2} \tag{18.7.7}$$

The diameter D, or the radius $a = D/2$, of the reflector and its focal length F determine the maximum angle ψ. It is obtained by setting $\rho = a$ in Eq. (18.7.4):

$$a = \frac{D}{2} = 2F\tan\left(\frac{\psi_0}{2}\right) \quad \Rightarrow \quad \boxed{\psi_0 = 2\,\mathrm{atan}\left(\frac{D}{4F}\right)} \tag{18.7.8}$$

Thus, the F/D ratio determines ψ_0. For example, if $F/D = 0.25, 0.35, 0.50$, then $\psi_0 = 90^{\circ}, 71^{\circ}, 53^{\circ}$. Practical F/D ratios are in the range 0.25–0.50.

18.8 Gain and Beamwidth of Reflector Antennas

To determine the radiation pattern of a reflector antenna, one may use Eq. (17.4.2), provided one knows the aperture fields $\mathbf{E}_a$, $\mathbf{H}_a$ on the effective aperture projected on the aperture plane. This approach is referred to as the *aperture-field method* [21].

Alternatively, the *current-distribution method* determines the current $\mathbf{J}_s$ on the surface of the reflector induced by the incident field from the feed, and then applies Eq. (17.4.1) with $\mathbf{J}_{ms} = 0$, using the curved surface of the reflector as the integration surface ($\mathbf{J}_{ms}$ vanishes on the reflector surface because there are no tangential electric fields on a perfect conductor.)

The two methods yield slightly different, but qualitatively similar, results for the radiation patterns. The aperture fields $\mathbf{E}_a, \mathbf{H}_a$ and the surface current $\mathbf{J}_s$ are determined by geometrical optics considerations based on the assumptions that (a) the reflector lies in the radiation zone of the feed antenna, and (b) the incident field from the feed gets reflected as if the reflector surface is perfectly conducting and locally flat. These assumptions are justified because in practice the size of the reflector and its curvature are much larger than the wavelength λ.

We use the polar and azimuthal angles ψ and χ indicated on Fig. 18.7.2 to characterize the direction $\hat{\mathbf{R}}$ of an incident ray from the feed to the reflector surface.

The radiated power from the feed within the solid angle $d\Omega = \sin\psi\,d\psi\,d\chi$ must be equal upon reflection to the power propagating parallel to the z-axis and intercepting the aperture plane through the area $dA = \rho\,d\rho\,d\chi$, as depicted in Fig. 18.7.1.

Assuming that $U_{\mathrm{feed}}(\psi,\chi)$ is the feed antenna's *radiation intensity* and noting that $|\mathbf{E}_a|^2/2\eta$ is the power density of the aperture field, the power condition reads:

$$\frac{1}{2\eta}\,|\mathbf{E}_a|^2 dA = U_{\mathrm{feed}}(\psi,\chi)\,d\Omega \quad \Rightarrow \quad \frac{1}{2\eta}\,|\mathbf{E}_a|^2\rho\,d\rho = U_{\mathrm{feed}}(\psi,\chi)\sin\psi\,d\psi \tag{18.8.1}$$

where we divided both sides by $d\chi$. Differentiating Eq. (18.7.4), we have:

$$d\rho = 2F\,\frac{d\psi}{2}\,\frac{1}{\cos^2(\psi/2)} = R\,d\psi$$

which implies that $\rho\,d\rho = R^2\sin\psi\,d\psi$. Thus, solving Eq. (18.8.1) for $|\mathbf{E}_a|$, we find:

$$\boxed{|\mathbf{E}_a(\rho,\chi)| = \frac{1}{R}\sqrt{2\eta U_{\mathrm{feed}}(\psi,\chi)}} \tag{18.8.2}$$

where we think of $\mathbf{E}_a$ as a function of $\rho = 2F\tan(\psi/2)$ and χ. Expressing R in terms of ρ, we have $R = 2F - h = F + (F - h) = F + \rho^2/4F$. Therefore, we may also write:

$$|\mathbf{E}_a(\rho,\chi)| = \frac{4F}{\rho^2 + 4F^2}\sqrt{2\eta U_{\mathrm{feed}}(\psi,\chi)} \tag{18.8.3}$$

Thus, the aperture fields get weaker towards the edge of the reflector. A measure of this tapering effect is the *edge illumination*, that is, the ratio of the electric field at the edge ($\rho = a$) and at the center ($\rho = 0$). Using Eqs. (18.7.3) and (18.8.2), we find:

$$\boxed{\frac{|\mathbf{E}_a(a,\chi)|}{|\mathbf{E}_a(0,\chi)|} = \frac{1+\cos\psi_0}{2}\sqrt{\frac{U_{\mathrm{feed}}(\psi_0,\chi)}{U_{\mathrm{feed}}(0,\chi)}}} \qquad \text{(edge illumination)} \tag{18.8.4}$$

In Sec. 17.6, we defined the directivity or gain of an aperture by Eq. (17.6.10), which we rewrite in the following form:

$$G_a = \frac{4\pi U_{\max}}{P_a} \tag{18.8.5}$$

where P_a is the total power through the aperture given in terms of $\mathbf{E}_a$ as follows:

$$P_a = \frac{1}{2\eta}\int_A |\mathbf{E}_a|^2 dA = \int_0^{\psi_0}\!\!\int_0^{2\pi} U_{\mathrm{feed}}(\psi,\chi)\sin\psi\,d\psi\,d\chi \tag{18.8.6}$$

and we used Eq. (18.8.1). For a reflector antenna, the gain must be defined relative to the total power P_{feed} of the feed antenna, that is,

$$G_{\mathrm{ant}} = \frac{4\pi U_{\max}}{P_{\mathrm{feed}}} = \frac{4\pi U_{\max}}{P_a}\,\frac{P_a}{P_{\mathrm{feed}}} = G_a e_{\mathrm{spl}} \tag{18.8.7}$$

The factor $e_{\mathrm{spl}} = P_a/P_{\mathrm{feed}}$ is referred to as the *spillover efficiency* or *loss* and represents the fraction of the power P_{feed} that actually gets reflected by the reflector antenna. The remaining power from the feed "spills over" the edge of the reflector and is lost.

We saw in Sec. 17.4 that the aperture gain is given in terms of the geometrical area A of the aperture and the aperture-taper and phase-error efficiencies by:

$$G_a = \frac{4\pi A}{\lambda^2}\,e_{\mathrm{atl}}\,e_{\mathrm{pel}} \tag{18.8.8}$$

It follows that the reflector antenna gain can be written as:

$$\boxed{G_{\mathrm{ant}} = G_a e_{\mathrm{spl}} = \frac{4\pi A}{\lambda^2}\,e_{\mathrm{atl}}\,e_{\mathrm{pel}}\,e_{\mathrm{spl}}} \tag{18.8.9}$$

The total aperture efficiency is $e_a = e_{\mathrm{atl}}\,e_{\mathrm{pel}}\,e_{\mathrm{spl}}$. In practice, additional efficiency or loss factors must be introduced, such as those due to cross polarization or to partial aperture blockage by the feed.

Of all the loss factors, the ATL and SPL are the primary ones that significantly affect the gain. Their tradeoff is captured by the *illumination efficiency* or *loss*, defined to be the product of ATL and SPL, $e_{\mathrm{ill}} = e_{\mathrm{atl}}\,e_{\mathrm{spl}}$.

The ATL and SPL may be expressed in terms of the radiation intensity $U_{\mathrm{feed}}(\psi,\chi)$. Using $\rho\,d\rho = R^2\sin\psi d\psi = \rho R d\psi = 2FR\tan(\psi/2)d\psi$ and Eq. (18.8.2), we have:

$$|\mathbf{E}_a|dA = \sqrt{2\eta U_{\mathrm{feed}}}\,\frac{1}{R}\,2FR\tan\frac{\psi}{2}\,d\psi\,d\chi = 2F\sqrt{2\eta U_{\mathrm{feed}}}\,\tan\frac{\psi}{2}\,d\psi\,d\chi$$

$$|\mathbf{E}_a|^2 dA = 2\eta U_{\mathrm{feed}}\,\frac{1}{R^2}\,R^2\sin\psi\,d\psi\,d\chi = 2\eta U_{\mathrm{feed}}\sin\psi\,d\psi\,d\chi$$

The aperture area is $A = \pi a^2 = \pi(2F)^2\tan^2(\psi_0/2)$. Thus, it follows from the definition (17.6.13) that the ATL will be:

$$e_{\text{atl}} = \frac{\left|\int_A |\mathbf{E}_a|\, dA\right|^2}{A\int_A |\mathbf{E}_a|^2 dA} = \frac{(2F)^2\left|\int_A \sqrt{2\eta U_{\text{feed}}}\,\tan\dfrac{\psi}{2}\,d\psi\,d\chi\right|^2}{\pi(2F)^2\tan^2(\psi_0/2)\int_A 2\eta U_{\text{feed}}\,\sin\psi\,d\psi\,d\chi}\,,\quad\text{or,}$$

$$e_{\text{atl}} = \frac{1}{\pi}\cot^2\left(\frac{\psi_0}{2}\right)\frac{\left|\int_0^{\psi_0}\int_0^{2\pi}\sqrt{U_{\text{feed}}(\psi,\chi)}\,\tan\dfrac{\psi}{2}\,d\psi\,d\chi\right|^2}{\int_0^{\psi_0}\int_0^{2\pi}U_{\text{feed}}(\psi,\chi)\sin\psi\,d\psi\,d\chi} \tag{18.8.10}$$

Similarly, the spillover efficiency can be expressed as:

$$e_{\text{spl}} = \frac{P_a}{P_{\text{feed}}} = \frac{\int_0^{\psi_0}\int_0^{2\pi}U_{\text{feed}}(\psi,\chi)\sin\psi\,d\psi\,d\chi}{\int_0^{\pi}\int_0^{2\pi}U_{\text{feed}}(\psi,\chi)\sin\psi\,d\psi\,d\chi} \tag{18.8.11}$$

where we replaced P_{feed} by the integral of U_{feed} over all solid angles. It follows that the illumination efficiency $e_{\text{ill}} = e_{\text{atl}}\,e_{\text{spl}}$ will be:

$$e_{\text{ill}} = \frac{1}{\pi}\cot^2\left(\frac{\psi_0}{2}\right)\frac{\left|\int_0^{\psi_0}\int_0^{2\pi}\sqrt{U_{\text{feed}}(\psi,\chi)}\,\tan\dfrac{\psi}{2}\,d\psi\,d\chi\right|^2}{\int_0^{\pi}\int_0^{2\pi}U_{\text{feed}}(\psi,\chi)\sin\psi\,d\psi\,d\chi} \tag{18.8.12}$$

An example of a feed pattern that approximates practical patterns is the following azimuthally symmetric radiation intensity [21]:

$$U_{\text{feed}}(\psi,\chi) = \begin{cases} U_0\cos^4\psi\,, & \text{if}\quad 0\le\psi\le\dfrac{\pi}{2} \\ 0\,, & \text{if}\quad \dfrac{\pi}{2}<\psi\le\pi \end{cases} \tag{18.8.13}$$

For this example, the SPL, ATL, and ILL can be computed in closed form:

$$\begin{aligned} e_{\text{spl}} &= 1-\cos^5\psi_0 \\ e_{\text{atl}} &= 40\cot^2(\psi_0/2)\,\frac{\left[\sin^4(\psi_0/2)+\ln\big(\cos(\psi_0/2)\big)\right]^2}{1-\cos^5\psi_0} \\ e_{\text{ill}} &= 40\cot^2(\psi_0/2)\,\left[\sin^4(\psi_0/2)+\ln\big(\cos(\psi_0/2)\big)\right]^2 \end{aligned} \tag{18.8.14}$$

The edge illumination is from Eq. (18.8.4):

$$\frac{|\mathbf{E}_a(\psi_0)|}{|\mathbf{E}_a(0)|} = \frac{1+\cos\psi_0}{2}\cos^2\psi_0 \tag{18.8.15}$$

Fig. 18.8.1 shows a plot of Eqs. (18.8.14) and (18.8.15) versus ψ_0. The ATL is a decreasing and the SPL an increasing function of ψ_0. The product $e_{\text{ill}} = e_{\text{atl}}\,e_{\text{spl}}$ reaches the maximum value of 0.82 at $\psi_0 = 53.31^\circ$. The corresponding edge illumination is 0.285 or −10.9 dB. The F/D ratio is $\cot(\psi_0/2)/4 = 0.498$.

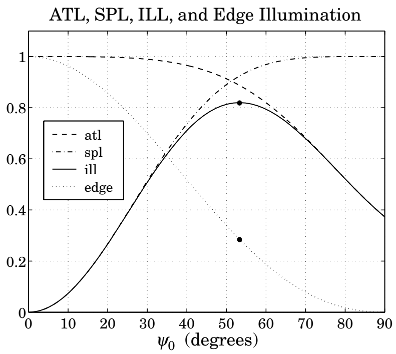


Fig. 18.8.1 Tradeoff between ATL and SPF.

This example gives rise to the rule of thumb that the best tradeoff between ATL and SPL for parabolic reflectors is achieved when the edge illumination is about −11 dB.

The value 0.82 for the efficiency is an overestimate. Taking into account other losses, the aperture efficiency of practical parabolic reflectors is typically of the order of 0.55–0.65. Expressing the physical area in terms of the diameter D, we can summarize the gain of a parabolic antenna:

$$\boxed{G = e_a\,\frac{4\pi A}{\lambda^2} = e_a\left(\frac{\pi D}{\lambda}\right)^2}\,,\quad\text{with}\quad e_a = 0.55\text{–}0.65 \tag{18.8.16}$$

As we discussed in Sec. 15.3, the 3-dB beamwidth of a reflector antenna with diameter D can be estimated by rule of thumb [1174]:

$$\boxed{\Delta\theta_{\text{3dB}} = 70^\circ\,\frac{\lambda}{D}} \tag{18.8.17}$$

The beamwidth depends also on the edge illumination. Typically, as the edge attenuation increases, the beamwidth widens and the sidelobes decrease. By studying various reflector sizes, types, and feeds Komen [1175] arrived at the following improved approximation for the 3-dB width, which takes into account the edge illumination:

$$\boxed{\Delta\theta_{\text{3dB}} = (1.05^\circ A_{\text{edge}} + 55.95^\circ)\,\frac{\lambda}{D}} \tag{18.8.18}$$

where A_{edge} is the edge attenuation in dB, that is, $A_{\text{edge}} = -20\log_{10}\big[|\mathbf{E}_a(\psi_0)/\mathbf{E}_a(0)|\big]$. For example, for $A_{\text{edge}} = 11$ dB, the angle factor becomes 67.5°.

18.9 Aperture-Field and Current-Distribution Methods

In the previous section, we used energy flow considerations to determine the magnitude $|\mathbf{E}_a|$ of the aperture field. To determine its direction and phase, we need to start from the field radiated by the feed antenna and trace its path as it propagates as a spherical wave to the reflector surface, gets reflected there, and then propagates as a plane wave along the z-direction to the aperture plane.

Points on the reflector surface will be parametrized by the spherical coordinates R, ψ, χ as shown in Figs. 18.7.1 and 18.7.2, and points in the radiation zone of the reflector antenna, by the usual r, θ, ϕ.

Let $\hat{\mathbf{R}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\chi}}$ be the unit vectors in the R, ψ, χ directions. The relationships of R, ψ, χ to the conventional polar coordinates of the $x'y'z'$ coordinate system are: $R = r'$, $\psi = \theta'$, but $\chi = -\phi'$, so that the unit vectors are $\hat{\mathbf{R}} = \hat{\mathbf{r}}'$, $\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\theta}}'$, and $\hat{\boldsymbol{\chi}} = -\hat{\boldsymbol{\phi}}'$. (The primed system has $\hat{\mathbf{x}}' = \hat{\mathbf{x}}$, $\hat{\mathbf{y}}' = -\hat{\mathbf{y}}$, and $\hat{\mathbf{x}}' = -\hat{\mathbf{z}}$.) In terms of the unprimed system:

$$
\begin{aligned}
\hat{\mathbf{R}} &= \hat{\mathbf{x}} \sin\psi\cos\chi + \hat{\mathbf{y}}\sin\psi\sin\chi - \hat{\mathbf{z}}\cos\psi \\
\hat{\boldsymbol{\psi}} &= \hat{\mathbf{x}} \cos\psi\cos\chi + \hat{\mathbf{y}}\cos\psi\sin\chi + \hat{\mathbf{z}}\sin\psi \\
\hat{\boldsymbol{\chi}} &= -\hat{\mathbf{x}}\sin\chi + \hat{\mathbf{y}}\cos\chi
\end{aligned}
\tag{18.9.1}
$$

and conversely,

$$
\begin{aligned}
\hat{\mathbf{x}} &= \hat{\mathbf{R}} \sin\psi\cos\chi + \hat{\boldsymbol{\psi}}\cos\psi\cos\chi - \hat{\boldsymbol{\chi}}\sin\chi \\
\hat{\mathbf{y}} &= \hat{\mathbf{R}} \sin\psi\sin\chi + \hat{\boldsymbol{\psi}}\cos\psi\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi \\
\hat{\mathbf{z}} &= -\hat{\mathbf{R}}\cos\psi + \hat{\boldsymbol{\psi}}\sin\psi
\end{aligned}
\tag{18.9.2}
$$

Because the reflector is assumed to be in the radiation zone of the feed, the most general field radiated by the feed, and incident at the point R, ψ, χ on the reflector surface, will have the form:

$$
\boxed{\boldsymbol{E}_i = \frac{e^{-jkR}}{R}\, \boldsymbol{f}_i(\psi, \chi)} \qquad \text{(incident field)} \tag{18.9.3}
$$

Because of the requirement $\hat{\mathbf{R}} \cdot \boldsymbol{E}_i = 0$, the vector function $\boldsymbol{f}_i$ must satisfy $\hat{\mathbf{R}} \cdot \boldsymbol{f}_i = 0$. As expected for radiation fields, the radial dependence on R is decoupled from the angular dependence on ψ, χ. The corresponding magnetic field will be:

$$
\boldsymbol{H}_i = \frac{1}{\eta}\hat{\mathbf{R}} \times \boldsymbol{E}_i = \frac{1}{\eta}\frac{e^{-jkR}}{R}\hat{\mathbf{R}} \times \boldsymbol{f}_i(\psi, \chi) \tag{18.9.4}
$$

The feed's radiation intensity U_{feed} is related to $\boldsymbol{f}_i$ through the definition:

$$
U_{\text{feed}}(\psi, \chi) = R^2 \frac{1}{2\eta}|\boldsymbol{E}_i|^2 = \frac{1}{2\eta}\big|\boldsymbol{f}_i(\psi, \chi)\big|^2 \tag{18.9.5}
$$

Assuming that the incident field is reflected locally like a plane wave from the reflector's perfectly conducting surface, it follows that the reflected fields $\boldsymbol{E}_r, \boldsymbol{H}_r$ must satisfy the following relationships, where where $\hat{\mathbf{n}}$ is the normal to the reflector:

$$
\begin{aligned}
\hat{\mathbf{n}} \times \boldsymbol{E}_r &= -\hat{\mathbf{n}} \times \boldsymbol{E}_i\,, \qquad & \hat{\mathbf{n}} \cdot \boldsymbol{E}_r &= \hat{\mathbf{n}} \cdot \boldsymbol{E}_i \\
\hat{\mathbf{n}} \times \boldsymbol{H}_r &= \hat{\mathbf{n}} \times \boldsymbol{H}_i\,, \qquad & \hat{\mathbf{n}} \cdot \boldsymbol{H}_r &= -\hat{\mathbf{n}} \cdot \boldsymbol{H}_i
\end{aligned}
\tag{18.9.6}
$$

These imply that $|\boldsymbol{E}_r| = |\boldsymbol{E}_i|$, $|\boldsymbol{H}_r| = |\boldsymbol{H}_i|$, and that:

$$
\boxed{
\begin{aligned}
\boldsymbol{E}_r &= -\boldsymbol{E}_i + 2\hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \boldsymbol{E}_i) \\
\boldsymbol{H}_r &= \boldsymbol{H}_i - 2\hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \boldsymbol{H}_i)
\end{aligned}
}
\tag{18.9.7}
$$

Thus, the net electric field $\boldsymbol{E}_i + \boldsymbol{E}_r$ is normal to the surface. Fig. 18.9.1 depicts these geometric relationships, assuming for simplicity that $\boldsymbol{E}_i$ is parallel to $\hat{\boldsymbol{\psi}}$.

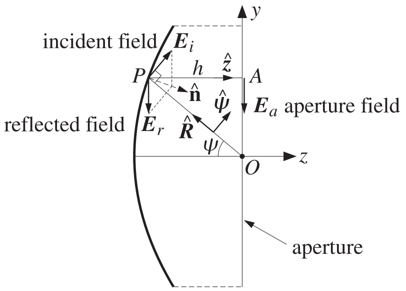


Fig. 18.9.1 Geometric relationship between incident and reflected electric fields.

The proof of Eq. (18.9.7) is straightforward. Indeed, using $\hat{\mathbf{n}} \times (\boldsymbol{E}_i + \boldsymbol{E}_r) = 0$ and the BAC-CAB rule, we have:

$$
0 = \big(\hat{\mathbf{n}} \times (\boldsymbol{E}_i + \boldsymbol{E}_r)\big) \times \hat{\mathbf{n}} = \boldsymbol{E}_i + \boldsymbol{E}_r - \hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \boldsymbol{E}_i + \hat{\mathbf{n}} \cdot \boldsymbol{E}_r) = \boldsymbol{E}_i + \boldsymbol{E}_r - \hat{\mathbf{n}}(2\,\hat{\mathbf{n}} \cdot \boldsymbol{E}_i)
$$

It follows now that the reflected field at the point (R, ψ, χ) will have the form:

$$
\boxed{\boldsymbol{E}_r = \frac{e^{-jkR}}{R}\, \boldsymbol{f}_r(\psi, \chi)} \qquad \text{(reflected field)} \tag{18.9.8}
$$

where $\boldsymbol{f}_r$ satisfies $|\boldsymbol{f}_r| = |\boldsymbol{f}_i|$ and:

$$
\boxed{\boldsymbol{f}_r = -\boldsymbol{f}_i + 2\hat{\mathbf{n}}(\hat{\mathbf{n}} \cdot \boldsymbol{f}_i)} \tag{18.9.9}
$$

The condition $\hat{\mathbf{R}} \cdot \boldsymbol{f}_i = 0$ implies that $\hat{\mathbf{z}} \cdot \boldsymbol{f}_r = 0$, so that $\boldsymbol{f}_r$ and $\boldsymbol{E}_r$ are perpendicular to the z-axis, and parallel to the aperture plane. To see this, we note that the normal $\hat{\mathbf{n}}$, bisecting the angle $\angle OPA$ in Fig. 18.9.1, will form an angle of $\psi/2$ with the z axis, so that $\hat{\mathbf{z}} \cdot \hat{\mathbf{n}} = \cos(\psi/2)$. More explicitly, the vector $\hat{\mathbf{n}}$ can be expressed in the form:

$$
\boxed{\hat{\mathbf{n}} = -\hat{\mathbf{R}}\cos\frac{\psi}{2} + \hat{\boldsymbol{\psi}}\sin\frac{\psi}{2} = \hat{\mathbf{z}}\cos\frac{\psi}{2} - (\hat{\mathbf{x}}\cos\chi + \hat{\mathbf{y}}\sin\chi)\sin\frac{\psi}{2}} \tag{18.9.10}
$$

Then, using Eq. (18.9.2), it follows that:

$$
\begin{aligned}
\hat{\mathbf{z}} \cdot \boldsymbol{f}_r &= -\hat{\mathbf{z}} \cdot \boldsymbol{f}_i + 2(\hat{\mathbf{z}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{n}} \cdot \boldsymbol{f}_i) \\
&= -(-\hat{\mathbf{R}}\cos\psi + \hat{\boldsymbol{\psi}}\sin\psi)\cdot\boldsymbol{f}_i + 2\cos\frac{\psi}{2}(-\hat{\mathbf{R}}\cos\frac{\psi}{2} + \hat{\boldsymbol{\psi}}\sin\frac{\psi}{2})\cdot\boldsymbol{f}_i \\
&= -(\hat{\boldsymbol{\psi}} \cdot \boldsymbol{f}_i)\left[\sin\psi - 2\cos\frac{\psi}{2}\sin\frac{\psi}{2}\right] = 0
\end{aligned}
$$

Next, we obtain the aperture field $\mathbf{E}_a$ by propagating $\mathbf{E}_r$ as a plane wave along the z-direction by a distance h to the aperture plane:

$$\mathbf{E}_a = e^{-jkh}\mathbf{E}_r = \frac{e^{-jk(R+h)}}{R}\,\mathbf{f}_r(\psi,\chi)$$

But for the parabola, we have $R + h = 2F$. Thus, the aperture field is given by:

$$\boxed{\mathbf{E}_a = \frac{e^{-2jkF}}{R}\,\mathbf{f}_a(\psi,\chi)} \qquad \text{(aperture field)} \tag{18.9.11}$$

where we defined $\mathbf{f}_a = \mathbf{f}_r$, so that:

$$\boxed{\mathbf{f}_a = -\mathbf{f}_i + 2\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{f}_i)} \tag{18.9.12}$$

Because $|\mathbf{f}_a| = |\mathbf{f}_r| = |\mathbf{f}_i| = \sqrt{2\eta U_{\text{feed}}}$, it follows that Eq. (18.9.11) is consistent with Eq. (18.8.2). As plane waves propagating in the z-direction, the reflected and aperture fields are Huygens sources. Therefore, the corresponding magnetic fields will be:

$$\mathbf{H}_r = \frac{1}{\eta}\hat{\mathbf{z}}\times\mathbf{E}_r\,, \quad \mathbf{H}_a = \frac{1}{\eta}\hat{\mathbf{z}}\times\mathbf{E}_a$$

The surface currents induced on the reflector are obtained by noting that the total fields are $\mathbf{E}_i + \mathbf{E}_r = 2\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{E}_i)$ and $\mathbf{H}_i + \mathbf{H}_r = 2\mathbf{H}_i - 2\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{H}_i)$. Thus, we have:

$$\begin{aligned}
\mathbf{J}_s &= \hat{\mathbf{n}}\times(\mathbf{H}_i + \mathbf{H}_r) = 2\,\hat{\mathbf{n}}\times\mathbf{H}_i = \frac{2}{\eta}\,\frac{e^{-jkR}}{R}\,\hat{\mathbf{R}}\times\mathbf{f}_i\\
\mathbf{J}_{ms} &= -\hat{\mathbf{n}}\times(\mathbf{E}_i + \mathbf{E}_r) = 0
\end{aligned}$$

18.10 Radiation Patterns of Reflector Antennas

The radiation patterns of the reflector antenna are obtained either from the aperture fields $\mathbf{E}_a, \mathbf{H}_a$ integrated over the effective aperture using Eq. (17.4.2), or from the currents $\mathbf{J}_s$ and $\mathbf{J}_{ms} = 0$ integrated over the curved reflector surface using Eq. (17.4.1).

We discuss in detail only the aperture-field case. The radiation fields at some large distance r in the direction defined by the polar angles θ, ϕ are given by Eq. (17.5.3). The unit vector $\hat{\mathbf{r}}$ in the direction of θ, ϕ is shown in Fig. 18.7.2. We have:

$$\begin{aligned}
E_\theta &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,\big[f_x\cos\phi + f_y\sin\phi\big]\\
E_\phi &= jk\,\frac{e^{-jkr}}{2\pi r}\,\frac{1+\cos\theta}{2}\,\big[f_y\cos\phi - f_x\sin\phi\big]
\end{aligned} \tag{18.10.1}$$

where the vector $\mathbf{f} = \hat{\mathbf{x}}f_x + \hat{\mathbf{y}}f_y$ is the Fourier transform over the aperture:

$$\mathbf{f}(\theta,\phi) = \int_0^a\int_0^{2\pi}\mathbf{E}_a(\rho',\chi)\,e^{j\mathbf{k}\cdot\mathbf{r}'}\rho'\,d\rho'\,d\chi \tag{18.10.2}$$

The vector $\mathbf{r}'$ lies on the aperture plane and is given in cylindrical coordinates by $\mathbf{r}' = \rho'\hat{\boldsymbol{\rho}} = \rho'(\hat{\mathbf{x}}\cos\chi + \hat{\mathbf{y}}\sin\chi)$. Thus,

$$\begin{aligned}
\mathbf{k}\cdot\mathbf{r}' &= k\rho'(\hat{\mathbf{x}}\cos\phi\sin\theta + \hat{\mathbf{y}}\sin\phi\sin\theta + \hat{\mathbf{z}}\cos\theta)\cdot(\hat{\mathbf{x}}\cos\chi + \hat{\mathbf{y}}\sin\chi)\\
&= k\rho'\sin\theta(\cos\phi\cos\chi + \sin\phi\sin\chi) = k\rho'\sin\theta\cos(\phi-\chi)
\end{aligned}$$

It follows that:

$$\boxed{\mathbf{f}(\theta,\phi) = \int_0^a\int_0^{2\pi}\mathbf{E}_a(\rho,\chi)\,e^{jk\rho\sin\theta\cos(\phi-\chi)}\rho\,d\rho\,d\chi} \tag{18.10.3}$$

We may convert this into an integral over the feed angles ψ, χ by using Eq. (18.9.11) and $d\rho = R\,d\psi$, $\rho = 2F\tan(\psi/2)$, and $\rho\,d\rho = 2FR\tan(\psi/2)\,d\psi$. Then, the $1/R$ factor in $\mathbf{E}_a$ is canceled, resulting in:

$$\boxed{\mathbf{f}(\theta,\phi) = 2Fe^{-2jkF}\int_0^{\psi_0}\int_0^{2\pi}\mathbf{f}_a(\psi,\chi)\,e^{2jkF\tan\frac{\psi}{2}\sin\theta\cos(\phi-\chi)}\tan\frac{\psi}{2}\,d\psi\,d\chi} \tag{18.10.4}$$

Given a feed pattern $\mathbf{f}_i(\psi,\chi)$, the aperture pattern $\mathbf{f}_a(\psi,\chi)$ is determined from Eq. (18.9.12) and the integrations in (18.10.4) are done numerically.

Because of the condition $\hat{\mathbf{R}}\cdot\mathbf{f}_i = 0$, the vector $\mathbf{f}_i$ will have components only along the $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\chi}}$ directions. We assume that $\mathbf{f}_i$ has the following more specific form:

$$\boxed{\mathbf{f}_i = \hat{\boldsymbol{\psi}}\,F_1\sin\chi + \hat{\boldsymbol{\chi}}\,F_2\cos\chi} \qquad \text{(y-polarized feeds)} \tag{18.10.5}$$

where F_1, F_2 are functions of ψ, χ, but often assumed to be functions only of ψ, representing the patterns along the principal planes $\chi = 90^{\text{o}}$ and $\chi = 0^{\text{o}}$.

Such feeds are referred to as "y-polarized" and include y-directed dipoles, and waveguides and horns in which the electric field on the horn aperture is polarized along the y direction (the x-polarized case is obtained by a rotation, replacing χ by $\chi + 90^{\text{o}}$.) Using Eqs. (18.9.1) and (18.9.10), the corresponding pattern $\mathbf{f}_a$ can be worked out:

$$\boxed{\mathbf{f}_a = -\hat{\mathbf{y}}\big[F_1\sin^2\chi + F_2\cos^2\chi\big] - \hat{\mathbf{x}}\big[(F_1 - F_2)\cos\chi\sin\chi\big]} \tag{18.10.6}$$

If $F_1 = F_2$, we have $\mathbf{f}_a = -\hat{\mathbf{y}}F_1$. But if $F_1 \neq F_2$, the aperture field $\mathbf{E}_a$ develops a "cross-polarized" component along the x direction. Various definitions of cross polarization have been discussed by Ludwig [1180].

As examples, we consider the cases of a y-directed Hertzian dipole feed, and waveguide and horn feeds. Adapting their radiation patterns given in Sections 16.2, 18.1, and 18.3, to the R, ψ, χ coordinate system, we obtain the following feed patterns, which are special cases of (18.10.5):

$$\begin{aligned}
\mathbf{f}_i(\psi,\chi) &= F_d\,(\hat{\boldsymbol{\psi}}\cos\psi\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi) && \text{(dipole feed)}\\
\mathbf{f}_i(\psi,\chi) &= F_w(\psi,\chi)\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi) && \text{(waveguide feed)}\\
\mathbf{f}_i(\psi,\chi) &= F_h(\psi,\chi)\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi) && \text{(horn feed)}
\end{aligned} \tag{18.10.7}$$

where F_d is the constant $F_d = -j\eta(Il)/2\lambda$, and F_w, F_h are given by:

$$
\begin{aligned}
F_w(\psi,\chi) &= -\frac{jabE_0}{\pi\lambda}(1+\cos\psi)\,\frac{\cos(\pi v_x)}{1-4v_x^2}\,\frac{\sin(\pi v_y)}{\pi v_y} \\
F_h(\psi,\chi) &= -\frac{jABE_0}{8\lambda}(1+\cos\psi)\,F_1(v_x,\sigma_a)\,F_0(v_y,\sigma_b)
\end{aligned}
\tag{18.10.8}
$$

where I, l are the current and length of the Hertzian dipole, a, b and A, B are the dimensions of the waveguide and horn apertures, and $v_x = (a/\lambda)\sin\psi\cos\chi$, $v_y = (b/\lambda)\sin\psi\sin\chi$ for the waveguide, and $v_x = (A/\lambda)\sin\psi\cos\chi$, $v_y = (B/\lambda)\sin\psi\sin\chi$, for the horn, and F_1, F_0 are the horn pattern functions defined in Sec. 18.3. The corresponding aperture patterns $\boldsymbol{f}_a$ are in the three cases:

$$
\begin{aligned}
\boldsymbol{f}_a(\psi,\chi) &= -\hat{\mathbf{y}}\,F_d\big[\cos\psi\sin^2\chi+\cos^2\chi\big]-\hat{\mathbf{x}}\,F_d\big[(\cos\psi-1)\sin\chi\cos\chi\big] \\
\boldsymbol{f}_a(\psi,\chi) &= -\hat{\mathbf{y}}\,F_w(\psi,\chi) \\
\boldsymbol{f}_a(\psi,\chi) &= -\hat{\mathbf{y}}\,F_h(\psi,\chi)
\end{aligned}
\tag{18.10.9}
$$

In the general case, a more convenient form of Eq. (18.10.6) is obtained by writing it in terms of the sum and difference patterns:

$$
A=\frac{F_1+F_2}{2}\,,\quad B=\frac{F_1-F_2}{2}\quad\Leftrightarrow\quad F_1=A+B\,,\quad F_2=A-B
\tag{18.10.10}
$$

Using some trigonometric identities, we may write (18.10.6) in the form:

$$
\boxed{\boldsymbol{f}_a=-\hat{\mathbf{y}}(A-B\cos 2\chi)-\hat{\mathbf{x}}(B\sin 2\chi)}
\tag{18.10.11}
$$

In general, A, B will be functions of ψ, χ (as in the waveguide and horn cases.) If we assume that they are functions only of ψ, then the χ-integration in the radiation pattern integral (18.10.4) can be done explicitly leaving an integral over ψ only. Using (18.10.11) and the Bessel-function identities,

$$
\int_0^{2\pi} e^{ju\cos(\phi-\chi)}\begin{bmatrix}\cos n\chi\\ \sin n\chi\end{bmatrix}d\chi = 2\pi j^n\begin{bmatrix}\cos n\phi\\ \sin n\phi\end{bmatrix}J_n(u)
\tag{18.10.12}
$$

we obtain:

$$
\boxed{\boldsymbol{f}(\theta,\phi)=-\hat{\mathbf{y}}\big[f_A(\theta)-f_B(\theta)\cos 2\phi\big]-\hat{\mathbf{x}}\big[f_B(\theta)\sin 2\phi\big]}
\tag{18.10.13}
$$

where the functions $f_A(\theta)$ and $f_B(\theta)$ are defined by:

$$
\begin{aligned}
f_A(\theta) &= \;\;\,4\pi F e^{-2jkF}\int_0^{\psi_0} A(\psi)\,J_0\Big(\frac{4\pi F}{\lambda}\tan\frac{\psi}{2}\sin\theta\Big)\tan\frac{\psi}{2}\,d\psi \\
f_B(\theta) &= -4\pi F e^{-2jkF}\int_0^{\psi_0} B(\psi)\,J_2\Big(\frac{4\pi F}{\lambda}\tan\frac{\psi}{2}\sin\theta\Big)\tan\frac{\psi}{2}\,d\psi
\end{aligned}
\tag{18.10.14}
$$

Using Eq. (18.10.13) and some trigonometric identities, we obtain:

$$
\begin{aligned}
f_x\cos\phi+f_y\sin\phi &= -(f_A+f_B)\sin\phi \\
f_y\cos\phi-f_x\sin\phi &= -(f_A-f_B)\cos\phi
\end{aligned}
$$

It follows that the radiation fields (18.10.1) are given by:

$$
\boxed{
\begin{aligned}
E_\theta &= -j\,\frac{e^{-jkr}}{\lambda r}\,\frac{1+\cos\theta}{2}\big[f_A(\theta)+f_B(\theta)\big]\sin\phi \\
E_\phi &= -j\,\frac{e^{-jkr}}{\lambda r}\,\frac{1+\cos\theta}{2}\big[f_A(\theta)-f_B(\theta)\big]\cos\phi
\end{aligned}}
\tag{18.10.15}
$$

Example 18.10.1: *Parabolic Reflector with Hertzian Dipole Feed.* We compute numerically the gain patterns for a y-directed Hertzian dipole feed. We take $F = 10\lambda$ and $D = 40\lambda$, so that $F/D = 0.25$ and $\psi_0 = 90^{\text{o}}$. These choices are similar to those in [1178].

Ignoring the constant F_d in (18.10.7), we have $F_1(\psi)= \cos\psi$ and $F_2(\psi)= 1$. Thus, the sum and difference patters are $A(\psi)= (\cos\psi + 1)/2$ and $B(\psi)= (\cos\psi - 1)/2$. Up to some overall constants, the required gain integrals will have the form:

$$
f_A(\theta)=\int_0^{\psi_0}F_A(\psi,\theta)\,d\psi\,,\qquad f_B(\theta)=\int_0^{\psi_0}F_B(\psi,\theta)\,d\psi
\tag{18.10.16}
$$

where

$$
\begin{aligned}
F_A(\psi,\theta) &= (1+\cos\psi)\,J_0\Big(\frac{4\pi F}{\lambda}\tan\frac{\psi}{2}\sin\theta\Big)\tan\frac{\psi}{2} \\
F_B(\psi,\theta) &= (1-\cos\psi)\,J_2\Big(\frac{4\pi F}{\lambda}\tan\frac{\psi}{2}\sin\theta\Big)\tan\frac{\psi}{2}
\end{aligned}
\tag{18.10.17}
$$

The integrals are evaluated numerically using Gauss-Legendre quadrature integration, which approximates an integral as a weighted sum [1267]:

$$
f_A(\theta)=\sum_{i=1}^{N}w_i\,F_A(\psi_i,\theta)=\mathbf{w}^T\mathbf{F}_A
$$

where w_i, ψ_i are the Gauss-Legendre weights and evaluation points within the integration interval $[0,\psi_0]$, where $\mathbf{F}_A$ is the column vector with ith component $F_A(\psi_i,\theta)$.

For higher accuracy, this interval may be subdivided into a number of subintervals, the quantities w_i, ψ_i are then determined on each subinterval, and the total integral is evaluated as the sum of the integrals over all the subintervals.

We have written a MATLAB function, `quadrs`, that determines the quantities w_i, ψ_i over all the subintervals. It is built on the function `quadr`, which determines the weights over a single interval.

The following MATLAB code evaluates and plots in Fig. 18.10.1 the E- and H-plane patterns (18.10.15) over the polar angles $0 \le \theta \le 5^{\text{o}}$.

```
F = 10; D = 40; psi0 = 2*acot(4*F/D);      % F/D = 0.25, ψ0 = 90o

ab = linspace(0, psi0, 5);                % 4 integration subintervals in [0, ψ0]
[w,psi] = quadrs(ab);                     % quadrature weights and evaluation points
                                          % uses 16 weights per subinterval
c = cos(psi); t = tan(psi/2);             % cos ψ, tan(ψ/2) at quadrature points

th = linspace(0, 5, 251);                     % angle θ in degrees over 0 ≤ θ ≤ 5o

for i=1:length(th),
    u = 4*pi*F*sin(th(i)*pi/180);             % u  =  2kF sin θ
```

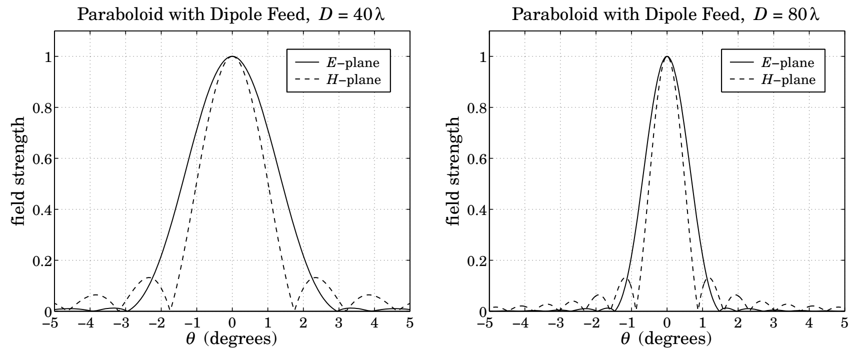


Fig. 18.10.1 Parabolic reflector patterns with dipole feed.

```
    FA = (1+c) .* besselj(0, u*t) .* t;         % integrand of fA(θ)
    fA(i) = w' * FA;                            % integral evaluated at θ
    FB = (1-c) .* besselj(2, u*t) .* t;         % integrand of fB(θ)
    fB(i) = w' * FB;
end

gh = abs((1+cos(th*pi/180)).*(fA-fB)); gh = gh/max(gh);      % gain patterns
ge = abs((1+cos(th*pi/180)).*(fA+fB)); ge = ge/max(ge);

plot(-th,ge,'-', th,ge, '-', -th,gh,'--',th,gh,'--');
```

The graph on the right has $\psi_0 = 90^\circ$ and $D = 80\lambda$, resulting in a narrower main beam. □

Example 18.10.2: *Parabolic Reflector with Waveguide Feed.* We calculate the reflector radiation patterns for a waveguide feed radiating in the TE$_{10}$ mode with a y-directed electric field. The feed pattern was given in Eq. (18.10.7). Ignoring some overall constants, we have with $v_x = (a/\lambda)\sin\psi\cos\chi$ and $v_y = (b/\lambda)\sin\psi\sin\chi$:

$$\boldsymbol{f}_i = (1+\cos\psi)\,\frac{\cos(\pi v_x)}{1-4v_x^2}\,\frac{\sin(\pi v_y)}{\pi v_y}\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi) \tag{18.10.18}$$

To avoid the double integration in the ψ and χ variables, we follow Jones' procedure [1178] of choosing the a, b such that the E- and H-plane illuminations of the paraboloid are essentially identical. This is accomplished when a is approximately $a = 1.37b$. Then, the above feed pattern may be simplified by replacing it by its E-plane pattern:

$$\boldsymbol{f}_i = (1+\cos\psi)\,\frac{\sin(\pi v_y)}{\pi v_y}\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi) \tag{18.10.19}$$

where $v_y = (b/\lambda)\sin\psi$. Thus, $F_1 = F_2$ and

$$A(\psi) = (1+\cos\psi)\,\frac{\sin(\pi b\sin\psi/\lambda)}{\pi b\sin\psi/\lambda} \quad\text{and}\quad B(\psi) = 0 \tag{18.10.20}$$

The radiated field is given by Eq. (18.10.15) with a normalized gain:

$$g(\theta) = \left|\frac{1+\cos\theta}{2}\,\frac{f_A(\theta)}{f_A(0)}\right|^2 \tag{18.10.21}$$

where $f_A(\theta)$ is defined up to a constant by Eq. (18.10.14):

$$f_A(\theta) = \int_0^{\psi_0} A(\psi)\,J_0\left(\frac{4\pi F}{\lambda}\tan\frac{\psi}{2}\sin\theta\right)\tan\frac{\psi}{2}\,d\psi \tag{18.10.22}$$

We choose a parabolic antenna with diameter $D = 40\lambda$ and subtended angle of $\psi_0 = 60^\circ$, so that $F = D\cot(\psi_0/2)/4 = 17.3205\lambda$. The length b of the waveguide is chosen such as to achieve an edge illumination of −11 dB on the paraboloid. This gives the condition on b, where the extra factor of $(1+\cos\psi)$ arises from the space attenuation factor $1/R$:

$$\frac{|\boldsymbol{E}_i(\psi_0)|}{|\boldsymbol{E}_i(0)|} = \left(\frac{1+\cos\psi_0}{2}\right)^2\left|\frac{\sin(\pi b\sin\psi_0/\lambda)}{\pi b\sin\psi_0/\lambda}\right| = 10^{-11/20} = 0.2818 \tag{18.10.23}$$

It has solution $b = 0.6958\lambda$ and therefore, $a = 1.37b = 0.9533\lambda$. The illumination efficiency given in Eq. (18.8.12) may be taken to be a measure of the overall aperture efficiency of the reflector. Because $2\eta U_{\text{feed}} = |\boldsymbol{f}_i|^2 = |\boldsymbol{f}_a|^2 = |A(\psi)|^2$, the integrals in (18.8.12) may be calculated numerically, giving $e_a = 0.71$ and a gain of 40.5 dB.

The pattern function $f_A(\theta)$ may be calculated numerically as in the previous example. The left graph in Fig. 18.10.2 shows the E- and H-plane illumination patterns versus ψ of the actual feed given by (18.10.18), that is, the normalized gains:

$$g_E(\psi) = \left|\frac{(1+\cos\psi)^2}{4}\,\frac{\sin(\pi b\sin\psi/\lambda)}{\pi b\sin\psi/\lambda}\right|^2$$

$$g_H(\psi) = \left|\frac{(1+\cos\psi)^2}{4}\,\frac{\cos(\pi a\sin\psi/\lambda)}{1-4(\pi a\sin\psi/\lambda)^2}\right|^2$$

They are essentially identical provided $a = 1.37b$ (the graph actually plots the square roots of these quantities.) The right graph shows the calculated radiation pattern $g(\theta)$ (or, rather its square root) of the paraboloid.

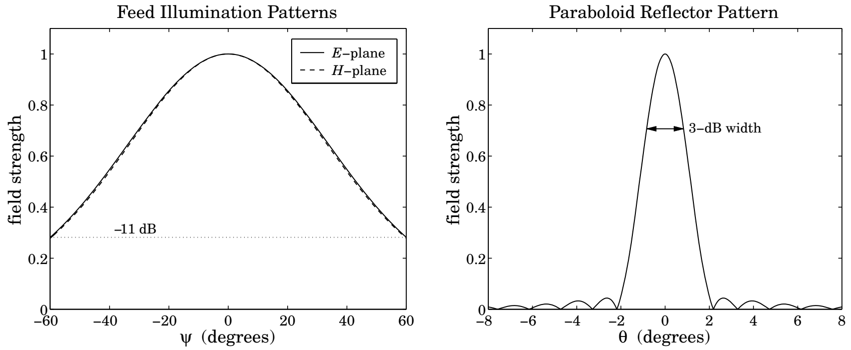


Fig. 18.10.2 Feed illumination and reflector radiation patterns.

The following MATLAB code solves (18.10.23) for b, and then calculates the illumination pattern and the reflector pattern:

```
F = 17.3205; D = 40; psi0 = 2*acot(4*F/D);           % ψ0 = 60°

f = inline('(1+cos(x)).^2/4 * abs(sinc(b*sin(x))) - A','b','x','A');
Aedge = 11;
b = fzero(f,0.8,optimset('display','off'), psi0, 10^(-Aedge/20));
a = 1.37 * b;

psi = linspace(-psi0, psi0, 201);  ps = psi * 180/pi;

gE = abs((1+cos(psi)).^2/4 .* sinc(b*sin(psi)));
gH = abs((1+cos(psi)).^2/4 .* dsinc(a*sin(psi)));

figure; plot(ps,gE,'-', ps,gH,'--');

[w,psi] = quadrs(linspace(0, psi0, 5));         % quadrature weights and points
s = sin(psi); c = cos(psi); t = tan(psi/2);
A = (1+c) .* sinc(b*s);                         % the pattern A(ψ)

thd = linspace(0, 5, 251); th = thd*pi/180;

for i=1:length(th),
    u = 4*pi*F*sin(th(i));
    FA = A .* besselj(0, u*t) .* t;
    fA(i) = w' * FA;
end

g = abs((1+cos(th)) .* fA); g = g/max(g);

figure; plot(-thd,g,'-', thd,g);
```

The 3-dB width was calculated from Eq. (18.8.18) and is placed on the graph. The angle factor was $1.05A_{\text{edge}} + 55.95 = 67.5$, so that $\Delta\theta_{3\text{dB}} = 67.5^{\circ}\lambda/D = 67.5/40 = 1.69^{\circ}$. The gain-beamwidth product is $p = G(\Delta\theta_{3\text{dB}})^2 = 10^{40.5/10}\,(1.69^{\circ})^2 = 32\,046\ \text{deg}^2$. □

Example 18.10.3: *Parabolic Reflector with Horn Feed.* Fig. 18.10.3 shows the illumination and reflector patterns if a rectangular horn antenna feed is used instead of a waveguide. The design requirements were again that the edge illumination be -11 dB and that $D = 40\lambda$ and $\psi_0 = 60^{\circ}$. The illumination pattern is (up to a scale factor):

$$\boldsymbol{f}_i = (1+\cos\psi)F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b)\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi)$$

The E- and H-plane illumination patterns are virtually identical over the angular range $0 \le \psi \le \psi_0$, provided one chooses the horn sides such that $A = 1.48B$. Then, the illumination field may be simplified by replacing it by the E-plane pattern and the length B is determined by requiring that the edge illumination be -11 dB. Therefore, we work with:

$$\boldsymbol{f}_i = (1+\cos\psi)F_0(\nu_y,\sigma_b)\,(\hat{\boldsymbol{\psi}}\sin\chi + \hat{\boldsymbol{\chi}}\cos\chi)\,, \quad \nu_y = \frac{B}{\lambda}\sin\psi$$

Then, $A(\psi) = (1+\cos\psi)F_0(\nu_y,\sigma_b)$ and $B(\psi) = 0$ for the sum and difference patterns. The edge illumination condition reads now:

$$\left(\frac{1+\cos\psi_0}{2}\right)^2 \left|\frac{F_0(\pi B\sin\psi_0/\lambda,\sigma_b)}{F_0(0,\sigma_b)}\right| = 10^{-11/20}$$

Its solution is $B = 0.7806\lambda$, and hence $A = 1.48B = 1.1553\lambda$. The left graph in Fig. 18.10.3 shows the E- and H-plane illumination gain patterns of the actual horn feed:

$$g_E(\psi) = \left|\frac{(1+\cos\psi)^2}{4}\,\frac{F_0(\pi B\sin\psi_0/\lambda,\sigma_b)}{F_0(0,\sigma_b)}\right|^2$$

$$g_H(\psi) = \left|\frac{(1+\cos\psi)^2}{4}\,\frac{F_1(\pi A\sin\psi_0/\lambda,\sigma_a)}{F_1(0,\sigma_a)}\right|^2$$

They are seen to be almost identical. The right graph shows the reflector radiation pattern computed numerically as in the previous example. The following MATLAB code illustrates this computation:

```
[w,psi] = quadrs(linspace(0, psi0, 5));          % 4 subintervals in [0, ψ0]

s = sin(psi); c = cos(psi); t = tan(psi/2);      % evaluate at quadrature points
Apsi = (1+c) .* (diffint(B*s, sb, 0));           % the pattern A(ψ)

thd = linspace(0, 8, 251); th = thd*pi/180;

for i=1:length(th),
    u = 4*pi*F*sin(th(i));
    FA = Apsi .* besselj(0, u*t) .* t;
    fA(i) = w' * FA;
end

g = abs((1+cos(th)) .* fA); g = g/max(g);

figure; plot(-thd,g,'-', thd,g);
```

The horn's σ-parameters were chosen to have the usual optimum values of $\sigma_a = 1.2593$ and $\sigma_b = 1.0246$. The 3-dB width is the same as in the previous example, that is, 1.69° and is shown on the graph. The computed antenna efficiency is now $e_a = 0.67$ and the corresponding gain 40.24 dB, so that $p = G(\Delta\theta_{3\text{dB}})^2 = 10^{40.24/10}\,(1.69^{\circ})^2 = 30\,184\ \text{deg}^2$ for the gain-beamwidth product. □

Example 18.10.4: Here, we compare the approximate symmetrized patterns of the previous two examples with the exact patterns obtained by performing the double-integration over the aperture variables ψ, χ.

Both the waveguide and horn examples have a y-directed two-dimensional Fourier transform pattern of the form:

$$f_A(\theta,\phi) = f_y(\theta,\phi) = \int_0^{\psi_0}\int_0^{2\pi} F_A(\psi,\chi,\theta,\phi)\,d\psi\,d\chi \tag{18.10.24}$$

where the integrand depends on the feed pattern $A(\psi,\chi)$:

$$F_A(\psi,\chi,\theta,\phi) = A(\psi,\chi)\,e^{j2kF\tan(\psi/2)\sin\theta\cos(\phi-\chi)}\tan\frac{\psi}{2} \tag{18.10.25}$$

and, up to constant factors, the function $A(\psi,\chi)$ is given in the two cases by:

$$\begin{aligned} A(\psi,\chi) &= (1+\cos\psi)\,\frac{\cos(\pi\nu_x)}{1-4\nu_x^2}\,\frac{\sin(\pi\nu_y)}{\pi\nu_y} \\ A(\psi,\chi) &= (1+\cos\psi)\,F_1(\nu_x,\sigma_a)\,F_0(\nu_y,\sigma_b) \end{aligned} \tag{18.10.26}$$

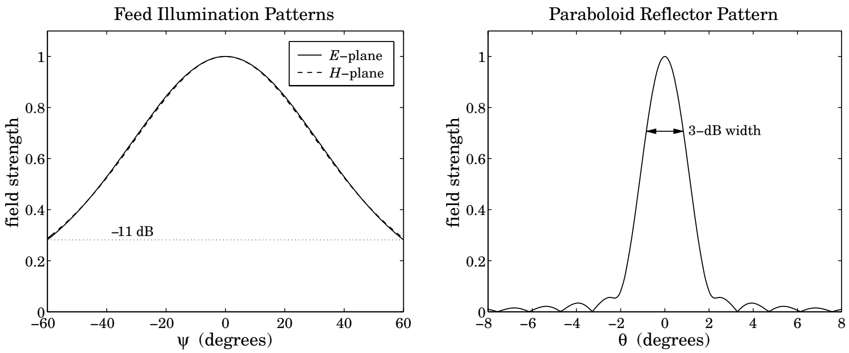


Fig. 18.10.3 Feed and reflector radiation patterns.

where $v_x = (a/\lambda)\sin\psi\cos\chi$ and $v_y = (b/\lambda)\sin\psi\sin\chi$ for the waveguide case, and $v_x = (A/\lambda)\sin\psi\cos\chi$ and $v_y = (B/\lambda)\sin\psi\sin\chi$ for the horn.

Once, $f_A(\theta,\phi)$ is computed, we obtain the (un-normalized) H- and E-plane radiation patterns for the reflector by setting $\phi = 0^{\text{o}}$ and 90^{o}, that is,

$$g_H(\theta)= \big|(1+\cos\theta)\,f_A(\theta,0^{\text{o}})\big|^2, \quad g_E(\theta)= \big|(1+\cos\theta)\,f_A(\theta,90^{\text{o}})\big|^2 \tag{18.10.27}$$

The numerical evaluation of Eq. (18.10.24) can be done with two-dimensional Gauss-Legendre quadratures, approximating the integral by the double sum:

$$f_A(\theta,\phi)=\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} w_{1i}\,F_A(\psi_i,\chi_j)\,w_{2j} = \mathbf{w}_1^T\mathbf{F}_A\mathbf{w}_2 \tag{18.10.28}$$

where $\{w_{1i},\psi_i\}$ and $\{w_{2j},\chi_j\}$ are the quadrature weights and evaluation points over the intervals $[0,\psi_0]$ and $[0,2\pi]$, and $\mathbf{F}_A$ is the matrix $F_A(\psi_i,\chi_j)$. The function `quadrs`, called on these two intervals, will generate these weights.

Fig. 18.10.4 shows the patterns (18.10.27) of the horn and waveguide cases evaluated numerically and plotted together with the approximate symmetrized patterns of the previous two examples. The symmetrized patterns agree very well with the exact patterns and fall between them. The following MATLAB code illustrates this computation for the horn case:

```
[w1, psi] = quadrs(linspace(0, psi0, Ni));    % quadrature over [0, ψ0], Ni = 5
[w2, chi] = quadrs(linspace(0, 2*pi, Ni));    % quadrature over [0, 2π], Ni = 5

sinpsi = sin(psi); cospsi = cos(psi); tanpsi = tan(psi/2);
sinchi = sin(chi); coschi = cos(chi);

for i = 1:length(chi),                        % build matrix A(ψi, χj) columnwise
   Apsi(:,i) = diffint(A*sinpsi*coschi(i), sa, 1) ...
            .* diffint(B*sinpsi*sinchi(i), sb, 0);
end
Apsi = repmat(tanpsi.*(1+cospsi), 1, length(psi)) .* Apsi;

th = linspace(0, 8, 401) * pi/180;
```

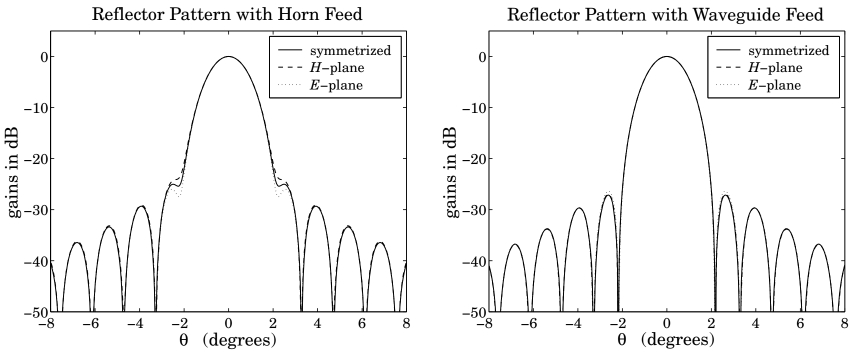


Fig. 18.10.4 Exact and approximate reflector radiation patterns.

```
for i=1:length(th),
   u = 4*pi*F*sin(th(i));                              % u = 2kF sin θ
   FH = Apsi .* exp(j*u*tanpsi*coschi');               % H-plane, φ = 0o
   FE = Apsi .* exp(j*u*tanpsi*sinchi');               % E-plane, φ = 90o
   fH(i) = w1' * FH * w2;                              % evaluate double integral
   fE(i) = w1' * FE * w2;
end

gH = abs((1+cos(th)).*fH); gH = gH/max(gH);            % radiation patterns
gE = abs((1+cos(th)).*fE); gE = gE/max(gE);
```

The patterns are plotted in dB, which accentuates the differences among the curves and also shows the sidelobe levels. In the waveguide case the resulting curves are almost indistinguishable to be seen as separate. ☐

18.11 Dual-Reflector Antennas

Dual-reflector antennas consisting of a main reflector and a secondary sub-reflector are used to increase the effective focal length and to provide convenient placement of the feed.

Fig. 18.11.1 shows a Cassegrain antenna[†] consisting of a parabolic reflector and a hyperbolic subreflector. The hyperbola is positioned such that its focus F_2 coincides with the focus of the parabola. The feed is placed at the other focus, F_1, of the hyperbola.

The focus F_2 is referred to a "virtual focus" of the parabola. Any ray originating from the point F_1 will be reflected by the hyperbola in a direction that appears to have originated from the focus F_2, and therefore, it will be re-reflected parallel to the parabola's axis.

To better understand the operation of such an antenna, we consider briefly the reflection properties of hyperbolas and ellipses, as shown in Fig. 18.11.2.

[†] Invented in the 17th century by A. Cassegrain.

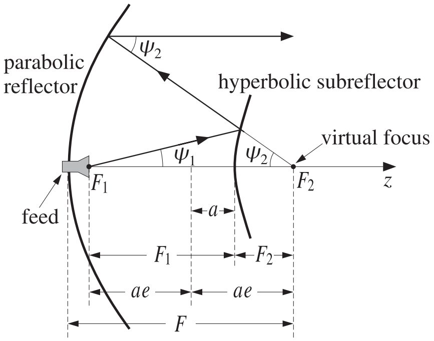


Fig. 18.11.1 Cassegrain dual-reflector antenna.

The geometrical properties of hyperbolas and ellipses are characterized completely by the parameters e, a, that is, the eccentricity and the distance of the vertices from the origin. The eccentricity is $e > 1$ for a hyperbola, and $e < 1$ for an ellipse. A circle corresponds to $e = 0$ and a parabola can be thought of as the limit of a hyperbola in the limit $e = 1$.

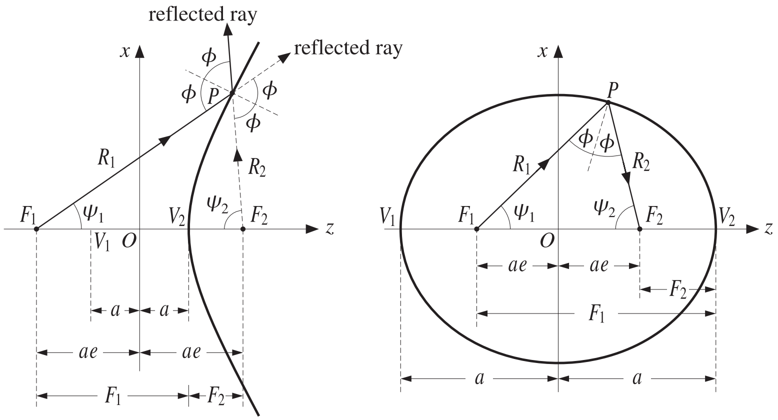


Fig. 18.11.2 Hyperbolic and elliptic reflectors.

The foci are at distances F_1 and F_2 from a vertex, say from the vertex V_2, and are given in terms of a, e as follows:

$$
\begin{aligned}
F_1 &= a(e+1), \quad F_2 = a(e-1) \qquad \text{(hyperbola)}\\
F_1 &= a(1+e), \quad F_2 = a(1-e) \qquad \text{(ellipse)}
\end{aligned}
\tag{18.11.1}
$$

The ray lengths R_1 and R_2 from the foci to a point P satisfy:

$$
\begin{aligned}
R_1 - R_2 &= 2a \qquad \text{(hyperbola)}\\
R_1 + R_2 &= 2a \qquad \text{(ellipse)}
\end{aligned}
\tag{18.11.2}
$$

The polar representations of the hyperbola or ellipse may be given in terms of the polar angles ψ_1 or ψ_2. We have:

$$
\begin{aligned}
R_1 &= \frac{a(e^2-1)}{e\cos\psi_1 - 1}, \quad R_2 = \frac{a(e^2-1)}{e\cos\psi_2 + 1} \qquad \text{(hyperbola)}\\
R_1 &= \frac{a(1-e^2)}{1 - e\cos\psi_1}, \quad R_2 = \frac{a(1-e^2)}{1 - e\cos\psi_2} \qquad \text{(ellipse)}
\end{aligned}
\tag{18.11.3}
$$

Note that we can write $a(e^2-1) = F_1(e-1) = F_2(e+1)$. For the hyperbola, the denominator of R_1 vanishes at the angles $\psi_1 = \pm\,\text{acos}(1/e)$, corresponding to two lines parallel to the hyperbola asymptotes.

In the cartesian coordinates x, z (defined with respect to the origin O in the figure), the equations for the hyperbola and the ellipse are:

$$
\begin{aligned}
(e^2-1)z^2 - x^2 &= a^2(e^2-1) \qquad \text{(hyperbola)}\\
(1-e^2)z^2 + x^2 &= a^2(1-e^2) \qquad \text{(ellipse)}
\end{aligned}
\tag{18.11.4}
$$

The semi-major axes are $b^2 = a^2(e^2-1)$ or $a^2(1-e^2)$. Because of the constraints (18.11.2), the angles ψ_1, ψ_2 are not independent of each other. For example, solving for ψ_2 in terms of ψ_1, we have in the hyperbolic case:

$$
\cos\psi_2 = \frac{e^2\cos\psi_1 - 2e + \cos\psi_1}{e^2 - 2e\cos\psi_1 + 1}
\tag{18.11.5}
$$

This implies the additional relationship and the derivative:

$$
\begin{aligned}
\frac{1+\cos\psi_2}{e\cos\psi_2 + 1} &= \left(\frac{1+\cos\psi_1}{e\cos\psi_1 - 1}\right)\left(\frac{e-1}{e+1}\right)\\
\frac{d\psi_2}{d\psi_1} &= \frac{\sin\psi_1}{\sin\psi_2}\left(\frac{e\cos\psi_2 + 1}{e\cos\psi_1 - 1}\right)^2
\end{aligned}
\tag{18.11.6}
$$

The incident ray R_1 reflects off the surface of either the hyperbola or the ellipse as though the surface is locally a perfect mirror, that is, the local normal bisects the angle between the incident and reflected rays. The angles of incidence and reflection ϕ shown on the figures are given by:

$$
\begin{aligned}
\phi &= \frac{\psi_1 + \psi_2}{2} \qquad \text{(hyperbola)}\\
\phi &= \frac{\pi}{2} - \frac{\psi_1 + \psi_2}{2} \qquad \text{(ellipse)}
\end{aligned}
\tag{18.11.7}
$$

To determine the aperture field on the aperture plane passing through F_2, we equate the power within a solid angle $d\Omega_1 = \sin\psi_1 d\psi_1 d\chi$ radiated from the feed, to the power reflected within the cone $d\Omega_2 = \sin\psi_2 d\psi_2 d\chi$ from the hyperbola, to the power passing through the aperture $dA = \rho d\rho d\chi$:

$$
dP = U_1(\psi_1,\chi)\,d\Omega_1 = U_2(\psi_2,\chi)\,d\Omega_2 = \frac{1}{2\eta}|\boldsymbol{E}_a|^2\,dA
\tag{18.11.8}
$$

where U_1 is the radiation intensity of the feed, and U_2 the intensity of the virtual feed. The second of Eqs. (18.11.8) may be solved as in Eq. (18.8.2) giving:

$$|\boldsymbol{E}_a| = \frac{1}{2F}\,(1+\cos\psi_2)\sqrt{2\eta U_2(\psi_2,\chi)} \tag{18.11.9}$$

where F is the focal length of the parabola. From the first of Eqs. (18.11.8), we find:

$$\sqrt{U_2} = \sqrt{U_1}\sqrt{\frac{\sin\psi_1 d\psi_1}{\sin\psi_2 d\psi_2}} = \sqrt{U_1}\,\frac{e\cos\psi_1 - 1}{e\cos\psi_2 + 1} \tag{18.11.10}$$

Inserting this into Eq. (18.11.9) and using Eqs. (18.11.6), we obtain:

$$|\boldsymbol{E}_a| = \frac{1}{2F}\left(\frac{e-1}{e+1}\right)(1+\cos\psi_1)\sqrt{2\eta U_1(\psi_1,\chi)} \tag{18.11.11}$$

Comparing with Eq. (18.8.2), we observe that this is equivalent to a *single* parabolic reflector with an effective focal length:

$$\boxed{F_{\text{eff}} = F\,\frac{e+1}{e-1}} \tag{18.11.12}$$

Thus, having a secondary reflector increases the focal length while providing a convenient location of the feed near the vertex of the parabola. Cassegrain antenna aperture efficiencies are typically of the order of 0.65–0.70.

18.12 Lens Antennas

Dielectric lens antennas convert the spherical wave from the feed into a plane wave exiting the lens. Fig. 18.12.1 shows two types of lenses, one with a hyperbolic and the other with elliptic profile.

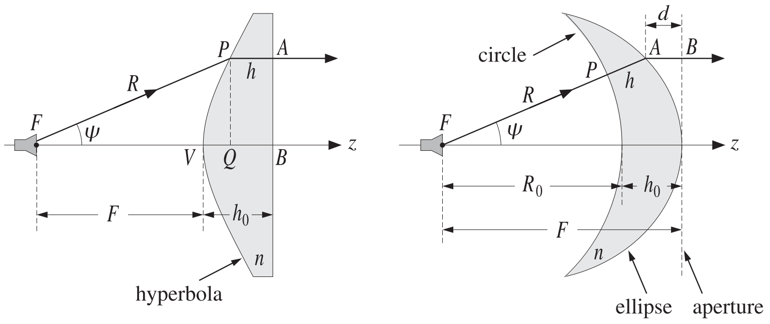


Fig. 18.12.1 Lens antennas.

The surface profile of the lens is determined by the requirement that the refracted rays all exit parallel to the lens axis. For example, for the lens shown on the left, the effective aperture plane is the right side AB of the lens. If this is to be the exiting wavefront, then each point A must have the same phase, that is, the same *optical path length* from the feed.

Taking the refractive index of the lens dielectric to be n, and denoting by R and h the lengths FP and PA, the constant-phase condition implies that the optical length along FPA be the same as that for FVB, that is,

$$R + nh = F + nh_0 \tag{18.12.1}$$

But, geometrically we have $R\cos\psi + h = F + h_0$. Multiplying this by n and subtracting Eq. (18.12.1), we obtain the polar equation for the lens profile:

$$R(n\cos\psi - 1) = F(n-1) \quad\Rightarrow\quad \boxed{R = \frac{F(n-1)}{n\cos\psi - 1}} \tag{18.12.2}$$

This is recognized from Eq. (18.11.3) to be the equation for a hyperbola with eccentricity and focal length $e = n$ and $F_1 = F$.

For the lens shown on the right, we assume the left surface is a circle of radius R_0 and we wish to determine the profile of the exiting surface such that the aperture plane is again a constant-phase wavefront. We denote by R and h the lengths FA and PA. Then, $R = R_0 + h$ and the constant-phase condition becomes:

$$R_0 + nh + d = R_0 + nh_0 \tag{18.12.3}$$

where the left-hand side represents the optical path $FPAB$. Geometrically, we have $R\cos\psi + d = F$ and $F = R_0 + h_0$. Eliminating d and R_0, we find the lens profile:

$$\boxed{R = \frac{F\left(1 - \dfrac{1}{n}\right)}{1 - \dfrac{1}{n}\cos\psi}} \tag{18.12.4}$$

which is recognized to be the equation for an ellipse with eccentricity and focal length $e = 1/n$ and $F_1 = F$.

In the above discussion, we considered only the refracted rays through the dielectric and ignored the reflected waves. These can be minimized by appropriate antireflection coatings.

19

Antenna Arrays

19.1 Antenna Arrays

Arrays of antennas are used to direct radiated power towards a desired angular sector. The number, geometrical arrangement, and relative amplitudes and phases of the array elements depend on the angular pattern that must be achieved.

Once an array has been designed to focus towards a particular direction, it becomes a simple matter to *steer* it towards some other direction by changing the relative phases of the array elements—a process called *steering* or *scanning*.

Figure 19.1.1 shows some examples of one- and two-dimensional arrays consisting of identical linear antennas. A linear antenna element, say along the z-direction, has an omnidirectional pattern with respect to the azimuthal angle ϕ. By replicating the antenna element along the x- or y-directions, the azimuthal symmetry is broken. By proper choice of the array feed coefficients a_n, any desired gain pattern $g(\phi)$ can be synthesized.

If the antenna element is replicated along the z-direction, then the omnidirectionality with respect to ϕ is maintained. With enough array elements, any prescribed polar angle pattern $g(\theta)$ can be designed.

In this section we discuss array design methods and consider various design issues, such as the tradeoff between beamwidth and sidelobe level.

For uniformly-spaced arrays, the design methods are identical to the methods for designing FIR digital filters in DSP, such as window-based and frequency-sampling designs. In fact, historically, these methods were first developed in antenna theory and only later were adopted and further developed in DSP.

19.2 Translational Phase Shift

The most basic property of an array is that the relative displacements of the antenna elements with respect to each other introduce relative *phase shifts* in the radiation vectors, which can then add constructively in some directions or destructively in others. This is a direct consequence of the translational phase-shift property of Fourier transforms: a translation in space or time becomes a phase shift in the Fourier domain.

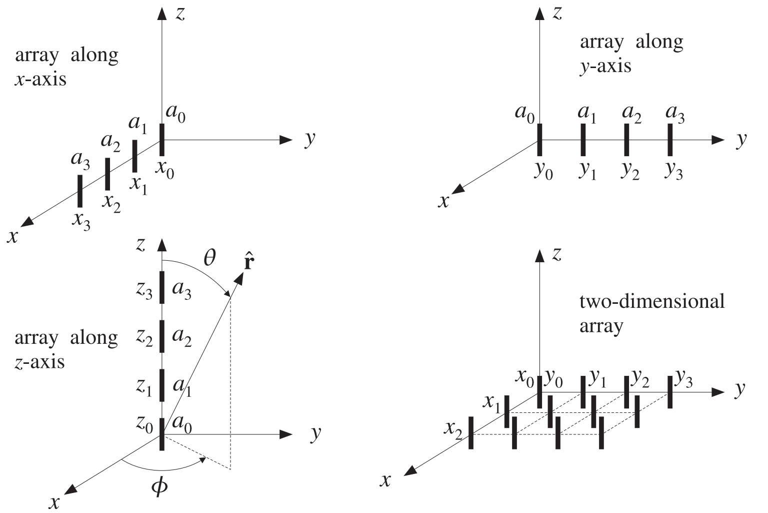


Fig. 19.1.1 Typical array configurations.

Figure 19.2.1 shows on the left an antenna translated by the vector $\mathbf{d}$, and on the right, several antennas translated to different locations and fed with different relative amplitudes.

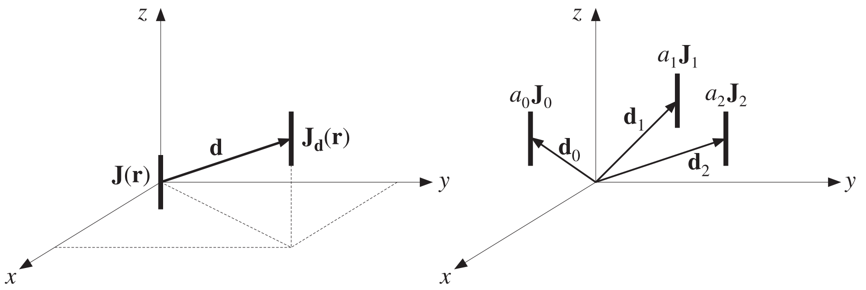


Fig. 19.2.1 Translated antennas.

The current density of the translated antenna will be $\boldsymbol{J}_{\mathbf{d}}(\mathbf{r})=\boldsymbol{J}(\mathbf{r}-\mathbf{d})$. By definition, the radiation vector is the three-dimensional Fourier transform of the current density, as in Eq. (14.7.5). Thus, the radiation vector of the translated current will be:

$$\begin{aligned}\boldsymbol{F}_{\mathbf{d}} &= \int e^{j\boldsymbol{k}\cdot\mathbf{r}}\,\boldsymbol{J}_{\mathbf{d}}(\mathbf{r})\,d^3\mathbf{r} = \int e^{j\boldsymbol{k}\cdot\mathbf{r}}\,\boldsymbol{J}(\mathbf{r}-\mathbf{d})\,d^3\mathbf{r} = \int e^{j\boldsymbol{k}\cdot(\mathbf{r}'+\mathbf{d})}\,\boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' \\ &= e^{j\boldsymbol{k}\cdot\mathbf{d}}\int e^{j\boldsymbol{k}\cdot\mathbf{r}'}\,\boldsymbol{J}(\mathbf{r}')\,d^3\mathbf{r}' = e^{j\boldsymbol{k}\cdot\mathbf{d}}\,\boldsymbol{F}\end{aligned}$$

where we changed variables to $\mathbf{r}'=\mathbf{r}-\mathbf{d}$. Thus,

$$\boxed{\boldsymbol{F}_{\mathbf{d}}(\boldsymbol{k})=e^{j\boldsymbol{k}\cdot\mathbf{d}}\,\boldsymbol{F}(\boldsymbol{k})}\qquad\text{(translational phase shift)}\tag{19.2.1}$$

19.3 Array Pattern Multiplication

More generally, we consider a three-dimensional array of several identical antennas located at positions $\mathbf{d}_0,\mathbf{d}_1,\mathbf{d}_2,\dots$ with relative feed coefficients $a_0,a_1,a_2,\dots$, as shown in Fig. 19.2.1. (Without loss of generality, we may set $\mathbf{d}_0=0$ and $a_0=1$.)

The current density of the nth antenna will be $\boldsymbol{J}_n(\mathbf{r})=a_n\boldsymbol{J}(\mathbf{r}-\mathbf{d}_n)$ and the corresponding radiation vector:

$$\boldsymbol{F}_n(\boldsymbol{k})=a_n e^{j\boldsymbol{k}\cdot\mathbf{d}_n}\,\boldsymbol{F}(\boldsymbol{k})$$

The total current density of the array will be:

$$\boldsymbol{J}_{\text{tot}}(\mathbf{r})=a_0\boldsymbol{J}(\mathbf{r}-\mathbf{d}_0)+a_1\boldsymbol{J}(\mathbf{r}-\mathbf{d}_1)+a_2\boldsymbol{J}(\mathbf{r}-\mathbf{d}_2)+\cdots$$

and the total radiation vector:

$$\boldsymbol{F}_{\text{tot}}(\boldsymbol{k})=\boldsymbol{F}_0+\boldsymbol{F}_1+\boldsymbol{F}_2+\cdots=a_0e^{j\boldsymbol{k}\cdot\mathbf{d}_0}\,\boldsymbol{F}(\boldsymbol{k})+a_1e^{j\boldsymbol{k}\cdot\mathbf{d}_1}\,\boldsymbol{F}(\boldsymbol{k})+a_2e^{j\boldsymbol{k}\cdot\mathbf{d}_2}\,\boldsymbol{F}(\boldsymbol{k})+\cdots$$

The factor $\boldsymbol{F}(\boldsymbol{k})$ due to a single antenna element at the origin is common to all terms. Thus, we obtain the *array pattern multiplication* property:

$$\boxed{\boldsymbol{F}_{\text{tot}}(\boldsymbol{k})=A(\boldsymbol{k})\boldsymbol{F}(\boldsymbol{k})}\qquad\text{(array pattern multiplication)}\tag{19.3.1}$$

where $A(\boldsymbol{k})$ is the *array factor*:

$$\boxed{A(\boldsymbol{k})=a_0e^{j\boldsymbol{k}\cdot\mathbf{d}_0}+a_1e^{j\boldsymbol{k}\cdot\mathbf{d}_1}+a_2e^{j\boldsymbol{k}\cdot\mathbf{d}_2}+\cdots}\qquad\text{(array factor)}\tag{19.3.2}$$

Since $\boldsymbol{k}=k\hat{\mathbf{r}}$, we may also denote the array factor as $A(\hat{\mathbf{r}})$ or $A(\theta,\phi)$. To summarize, the net effect of an array of identical antennas is to modify the single-antenna radiation vector by the array factor, which incorporates all the translational phase shifts and relative weighting coefficients of the array elements.

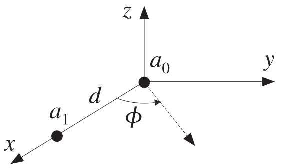
We may think of Eq. (19.3.1) as the input/output equation of a linear system with $A(\boldsymbol{k})$ as the *transfer function*. We note that the corresponding radiation intensities and power gains will also be related in a similar fashion:

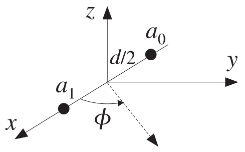
$$\boxed{\begin{aligned}U_{\text{tot}}(\theta,\phi)&=|A(\theta,\phi)|^2\,U(\theta,\phi)\\ G_{\text{tot}}(\theta,\phi)&=|A(\theta,\phi)|^2\,G(\theta,\phi)\end{aligned}}\tag{19.3.3}$$

where $U(\theta,\phi)$ and $G(\theta,\phi)$ are the radiation intensity and power gain of a single element. The array factor can dramatically alter the directivity properties of the single-antenna element. The power gain $|A(\theta,\phi)|^2$ of an array can be computed with the help of the MATLAB function `gain1d` of Appendix I with typical usage:

```
[g, phi] = gain1d(d, a, Nph);          % compute normalized gain of an array
```

Example 19.3.1: Consider an array of two isotropic antennas at positions $\mathbf{d}_0 = 0$ and $\mathbf{d}_1 = \hat{\mathbf{x}}d$ (alternatively, at $\mathbf{d}_0 = -(d/2)\hat{\mathbf{x}}$ and $\mathbf{d}_1 = (d/2)\hat{\mathbf{x}}$), as shown below:





The displacement phase factors are:

$$e^{j\mathbf{k}\cdot\mathbf{d}_0} = 1\,, \qquad e^{j\mathbf{k}\cdot\mathbf{d}_1} = e^{jk_x d} = e^{jkd\sin\theta\cos\phi}$$

or, in the symmetric case:

$$e^{j\mathbf{k}\cdot\mathbf{d}_0} = e^{-jk_x d/2} = e^{-jk(d/2)\sin\theta\cos\phi}\,, \qquad e^{j\mathbf{k}\cdot\mathbf{d}_1} = e^{jk_x d/2} = e^{jk(d/2)\sin\theta\cos\phi}$$

Let $\mathbf{a} = [a_0, a_1]$ be the array coefficients. The array factor is:

$$\begin{aligned} A(\theta,\phi) &= a_0 + a_1 e^{jkd\sin\theta\cos\phi} \\ A(\theta,\phi) &= a_0 e^{-jk(d/2)\sin\theta\cos\phi} + a_1 e^{jk(d/2)\sin\theta\cos\phi}\,, \qquad \text{(symmetric case)} \end{aligned}$$

The two expressions differ by a phase factor, which does not affect the power pattern. At polar angle $\theta = 90^{\circ}$, that is, on the xy-plane, the array factor will be:

$$A(\phi) = a_0 + a_1 e^{jkd\cos\phi}$$

and the azimuthal power pattern:

$$g(\phi) = |A(\phi)|^2 = \left|a_0 + a_1 e^{jkd\cos\phi}\right|^2$$

Note that $kd = 2\pi d/\lambda$. Figure 19.3.1 shows $g(\phi)$ for the array spacings $d = 0.25\lambda$, $d = 0.50\lambda$, $d = \lambda$, or $kd = \pi/2, \pi, 2\pi$, and the following array weights:

$$\begin{aligned} \mathbf{a} &= [a_0, a_1] = [1, 1] \\ \mathbf{a} &= [a_0, a_1] = [1, -1] \\ \mathbf{a} &= [a_0, a_1] = [1, -j] \end{aligned} \tag{19.3.4}$$

The first of these graphs was generated by the MATLAB code:

```
d = 0.25; a = [1,1];                    % d is in units of λ
[g, phi] = gain1d(d, a, 400);           % 400 phi's in [0,π]
dbz(phi, g, 30, 20);                    % 30° grid, 20-dB scale
```

As the relative phase of a_0 and a_1 changes, the pattern rotates so that its main lobe is in a different direction. When the coefficients are in phase, the pattern is broadside to the array, that is, towards $\phi = 90^{\circ}$. When they are in anti-phase, the pattern is end-fire, that is, towards $\phi = 0^{\circ}$ and $\phi = 180^{\circ}$.

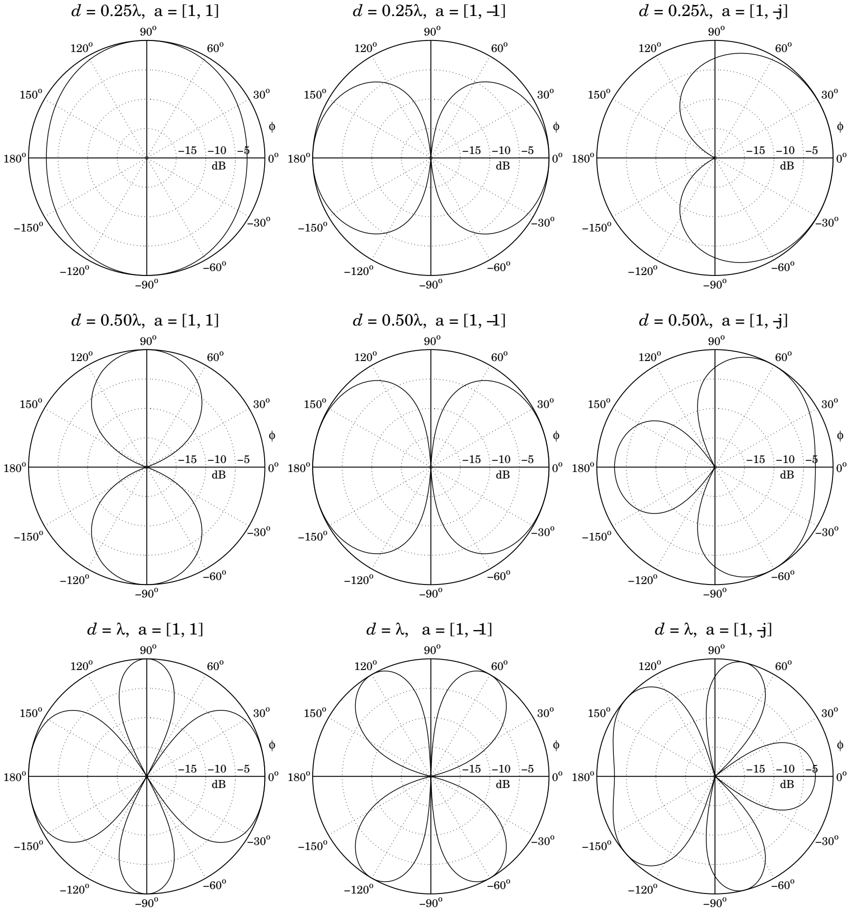


Fig. 19.3.1 Azimuthal gain patterns of two-element isotropic array.

The technique of rotating or steering the pattern towards some other direction by introducing relative phases among the elements is further discussed in Sec. 19.9. There, we will be able to predict the steering angles of this example from the relative phases of the weights.

Another observation from these graphs is that as the array pattern is steered from broadside to endfire, the widths of the main lobes become larger. We will discuss this effect in Sects. 19.9 and 19.10.

When $d \geq \lambda$, more than one main lobes appear in the pattern. Such main lobes are called grating lobes or fringes and are further discussed in Sec. 19.6. Fig. 19.3.2 shows some additional examples of grating lobes for spacings $d = 2\lambda$, 4λ, and 8λ. □

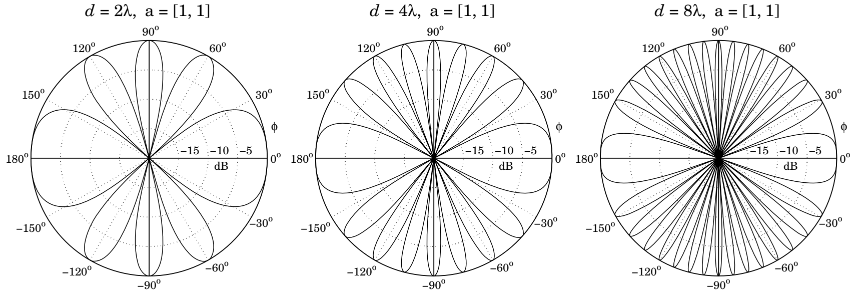
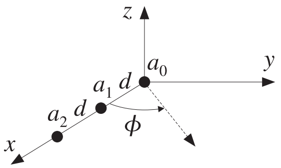
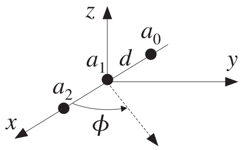

$d = 2\lambda$, a = [1, 1] $d = 4\lambda$, a = [1, 1] $d = 8\lambda$, a = [1, 1]

Fig. 19.3.2 Grating lobes of two-element isotropic array.

Example 19.3.2: Consider a three-element array of isotropic antennas at locations $\mathbf{d}_0 = 0$, $\mathbf{d}_1 = d\hat{\mathbf{x}}$, and $\mathbf{d}_2 = 2d\hat{\mathbf{x}}$, or, placed symmetrically at $\mathbf{d}_0 = -d\hat{\mathbf{x}}$, $\mathbf{d}_1 = 0$, and $\mathbf{d}_2 = d\hat{\mathbf{x}}$, as shown below:





The displacement phase factors evaluated at $\theta = 90^{\circ}$ are:

$$e^{j\mathbf{k}\cdot\mathbf{d}_0} = 1\,,\quad e^{j\mathbf{k}\cdot\mathbf{d}_1} = e^{jk_x d} = e^{jkd\cos\phi}\quad e^{j\mathbf{k}\cdot\mathbf{d}_2} = e^{j2k_x d} = e^{j2kd\cos\phi}$$

Let $\mathbf{a} = [a_0, a_1, a_2]$ be the array weights. The array factor is:

$$A(\phi) = a_0 + a_1 e^{jkd\cos\phi} + a_2 e^{2jkd\cos\phi}$$

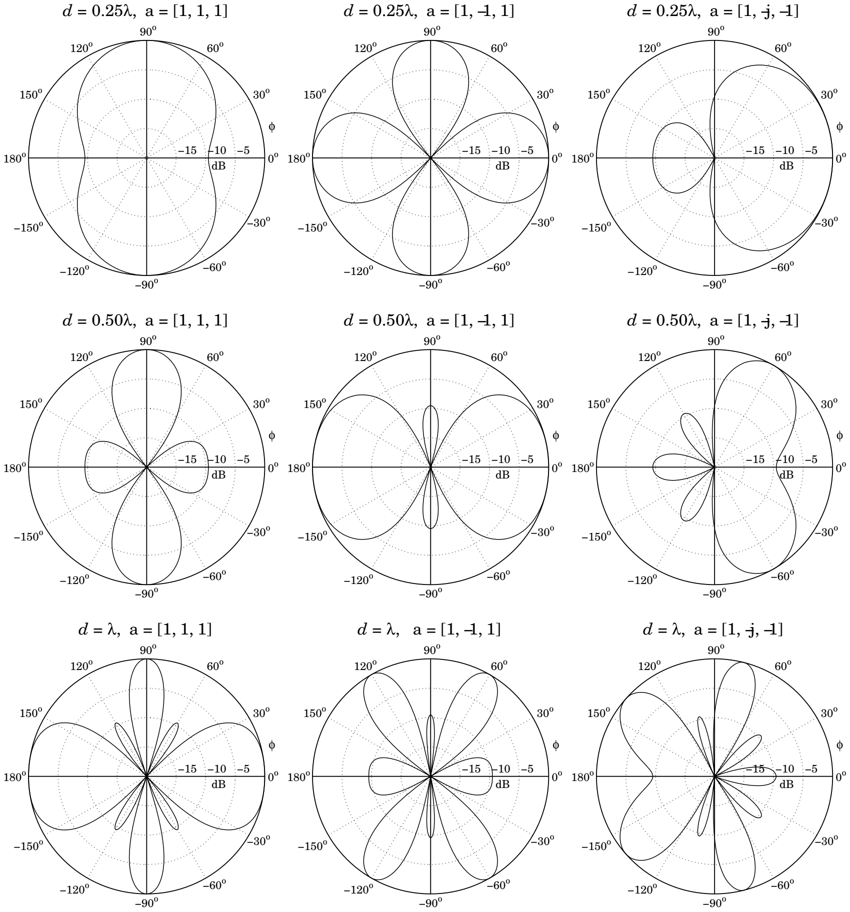
Figure 19.3.3 shows $g(\phi) = |A(\phi)|^2$ for the array spacings $d = 0.25\lambda$, $d = 0.50\lambda$, $d = \lambda$, or $kd = \pi/2, \pi, 2\pi$, and the following choices for the weights:

$$\begin{aligned}
\mathbf{a} &= [a_0, a_1, a_2] = [1, 1, 1] \\
\mathbf{a} &= [a_0, a_1, a_2] = [1, (-1), (-1)^2] = [1, -1, 1] \qquad (19.3.5)\\
\mathbf{a} &= [a_0, a_1, a_2] = [1, (-j), (-j)^2] = [1, -j, -1]
\end{aligned}$$

where in the last two cases, progressive phase factors of 180° and 90° have been introduced between the array elements.

The MATLAB code for generating the last graph was:

```
d = 1; a = [1,-j,-1];
[g, phi] = gain1d(d, a, 400);
dbz(phi, g, 30, 20);
```


$d = 0.25\lambda$, a = [1, 1, 1] $d = 0.25\lambda$, a = [1, −1, 1] $d = 0.25\lambda$, a = [1, −j, −1]

$d = 0.50\lambda$, a = [1, 1, 1] $d = 0.50\lambda$, a = [1, −1, 1] $d = 0.50\lambda$, a = [1, −j, −1]

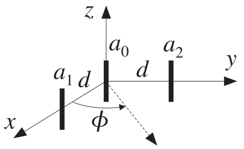
$d = \lambda$, a = [1, 1, 1] $d = \lambda$, a = [1, −1, 1] $d = \lambda$, a = [1, −j, −1]

Fig. 19.3.3 Azimuthal gains of three-element isotropic array.

The patterns are similarly rotated as in the previous example. The main lobes are narrower, but we note the appearance of sidelobes at the level of −10 dB. We will see later that as the number of array elements increases, the sidelobes reach a constant level of about −13 dB for an array with uniform weights.

Such sidelobes can be reduced further if we use appropriate non-uniform weights, but at the expense of increasing the beamwidth of the main lobes. □

Example 19.3.3: As an example of a two-dimensional array, consider three z-directed half-wave dipoles: one at the origin, one on the x-axis, and one on the y-axis, both at a distance $d = \lambda/2$, as shown below.



The relative weights are a_0, a_1, a_2. The displacement vectors are $\mathbf{d}_1 = \hat{\mathbf{x}}d$ and $\mathbf{d}_2 = \hat{\mathbf{y}}d$. Using Eq. (16.1.4), we find the translational phase-shift factors:

$$e^{j\boldsymbol{k}\cdot\mathbf{d}_1} = e^{jk_x d} = e^{jkd\sin\theta\cos\phi}\,, \qquad e^{j\boldsymbol{k}\cdot\mathbf{d}_2} = e^{jk_y d} = e^{jkd\sin\theta\sin\phi}$$

and the array factor:

$$A(\theta,\phi) = a_0 + a_1 e^{jkd\sin\theta\cos\phi} + a_2 e^{jkd\sin\theta\sin\phi}$$

Thus, the array's total normalized gain will be up to an overall constant:

$$g_{\text{tot}}(\theta,\phi) = |A(\theta,\phi)|^2\, g(\theta,\phi) = |A(\theta,\phi)|^2 \left|\frac{\cos(0.5\pi\cos\theta)}{\sin\theta}\right|^2$$

The gain pattern on the xy-plane ($\theta = 90^{\text{o}}$) becomes:

$$g_{\text{tot}}(\phi) = \left|a_0 + a_1 e^{jkd\cos\phi} + a_2 e^{jkd\sin\phi}\right|^2$$

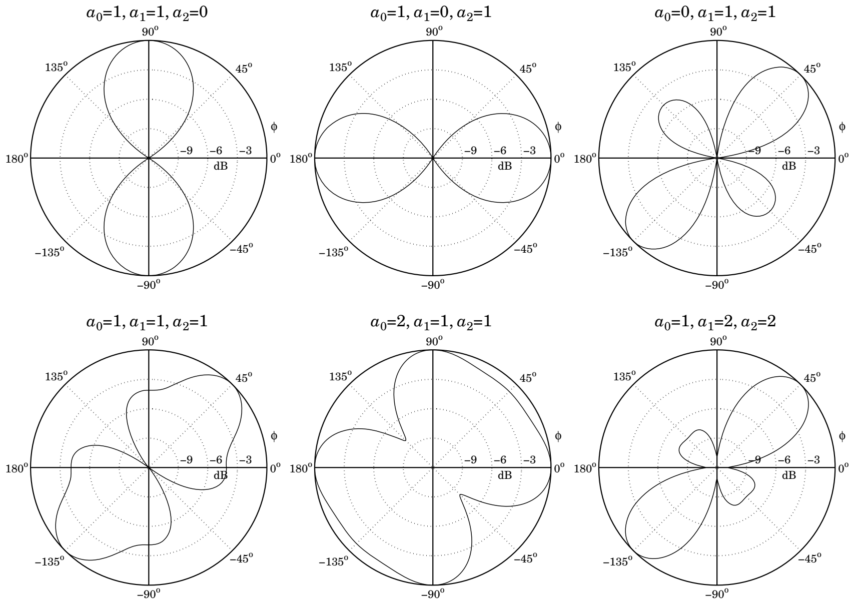
Note that because $d = \lambda/2$, we have $kd = \pi$. The omnidirectional case of a single element is obtained by setting $a_1 = a_2 = 0$ and $a_0 = 1$. Fig. 19.3.4 shows the gain $g_{\text{tot}}(\phi)$ for various choices of the array weights a_0, a_1, a_2.

Because of the presence of the a_2 term, which depends on $\sin\phi$, the gain is not necessarily symmetric for negative ϕ's. Thus, it must be evaluated over the entire azimuthal range $-\pi \le \phi \le \pi$. Then, it can be plotted with the help of the function `dbz2` which assumes the gain is over the entire 2π range. For example, the last of these graphs was computed by:

```
d = 0.5; a0=1; a1=2; a2=2;
phi = (0:400) * 2*pi/400;
psi1 = 2*pi*d*cos(phi);
psi2 = 2*pi*d*sin(phi);
g = abs(a0 + a1 * exp(j*psi1) + a2 * exp(j*psi2)).^2;
g = g/max(g);
dbz2(phi, g, 45, 12);
```

When $a_2 = 0$, we have effectively a two-element array along the x-axis with equal weights. The resulting array pattern is broadside, that is, maximum along the perpendicular $\phi = 90^{\text{o}}$ to the array. Similarly, when $a_1 = 0$, the two-element array is along the y-axis and the pattern is broadside to it, that is, along $\phi = 0$. When $a_0 = 0$, the pattern is broadside to the line joining elements 1 and 2. □

Example 19.3.4: The analysis of the rhombic antenna in Sec. 16.7 was carried out with the help of the translational phase-shift theorem of Eq. (19.2.1). The theorem was applied to antenna pairs 1, 3 and 2, 4.


a_0=1, a_1=1, a_2=0 a_0=1, a_1=0, a_2=1 a_0=0, a_1=1, a_2=1

a_0=1, a_1=1, a_2=1 a_0=2, a_1=1, a_2=1 a_0=1, a_1=2, a_2=2

Fig. 19.3.4 Azimuthal gain patterns of two-dimensional array.

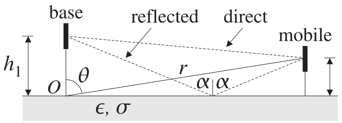
A more general version of the translation theorem involves both a translation and a rotation (a Euclidean transformation) of the type $\mathbf{r}' = R^{-1}(\mathbf{r} - \mathbf{d})$, or, $\mathbf{r} = R\mathbf{r}' + \mathbf{d}$, where R is a rotation matrix.

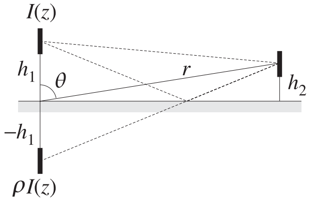
The rotated/translated current density is then defined as $\boldsymbol{J}_{R,\mathbf{d}}(\mathbf{r}) = R^{-1}\boldsymbol{J}(\mathbf{r}')$ and the corresponding relationship between the two radiation vectors becomes:

$$\boldsymbol{F}_{R,\mathbf{d}}(\boldsymbol{k}) = e^{j\boldsymbol{k}\cdot\mathbf{d}} R^{-1}\boldsymbol{F}(R^{-1}\boldsymbol{k})$$

The rhombic as well as the vee antennas can be analyzed by applying such rotational and translational transformations to a single traveling-wave antenna along the z-direction, which is rotated by an angle $\pm\alpha$ and then translated. □

Example 19.3.5: *Ground Effects Between Two Antennas.* There is a large literature on radiowave propagation effects [19,34,43,1190–1206]. Consider a mobile radio channel in which the transmitting vertical antenna at the base station is at height h_1 from the ground and the receiving mobile antenna is at height h_2, as shown below. The ray reflected from the ground interferes with the direct ray and can cause substantial signal cancellation at the receiving antenna.





The reflected ray may be thought of as originating from the *image* of the transmitting antenna at $-h_1$, as shown. Thus, we have an equivalent two-element transmitting array. We assume that the currents on the actual and image antennas are $I(z)$ and $\rho I(z)$, where $\rho = -\rho_{TM}$ is the reflection coefficient of the ground for parallel polarization (the negative sign is justified in the next example), given in terms of the angle of incidence α by:

$$\rho = -\rho_{TM} = \frac{n^2\cos\alpha - \sqrt{n^2 - \sin^2\alpha}}{n^2\cos\alpha + \sqrt{n^2 - \sin^2\alpha}}\,, \qquad n^2 = \frac{\epsilon}{\epsilon_0} - j\frac{\sigma}{\omega\epsilon_0} = \epsilon_r - j\frac{\eta_0}{2\pi}\sigma\lambda$$

where n is the complex refractive index of the ground, and we replaced $\omega\epsilon_0 = 2\pi f\epsilon_0 = 2\pi c_0\epsilon_0/\lambda$ and $c_0\epsilon_0 = 1/\eta_0$. Numerically, we may set $\eta_0/2\pi \simeq 60\ \Omega$. From the geometry of the figure, we find that the angle α is related to the polar angle θ by:

$$\tan\alpha = \frac{r\sin\theta}{h_1 + r\cos\theta}$$

In the limit of large r, α tends to θ. For a perfectly conducting ground ($\sigma = \infty$), the reflection coefficient becomes $\rho = 1$, regardless of the incidence angle.

On the other hand, for an imperfect ground and for *low* grazing angles ($\alpha \simeq 90^{\mathrm{o}}$), the reflection coefficient becomes $\rho = -1$, regardless of the conductivity of the ground. This is the relevant case in mobile communications.

The array factor can be obtained as follows. The two displaced antennas are at locations $\mathbf{d}_1 = h_1\hat{\mathbf{z}}$ and $\mathbf{d}_2 = -h_1\hat{\mathbf{z}}$, so that the displacement phase factors are:

$$e^{j\mathbf{k}\cdot\mathbf{d}_1} = e^{jk_z h_1} = e^{jkh_1\cos\theta}\,, \qquad e^{j\mathbf{k}\cdot\mathbf{d}_2} = e^{-jk_z h_1} = e^{-jkh_1\cos\theta}$$

where we replaced $k_z = k\cos\theta$. The relative feed coefficients are 1 and ρ. Therefore, the array factor and its magnitude will be:

$$\begin{aligned} A(\theta) &= e^{jkh_1\cos\theta} + \rho\, e^{-jkh_1\cos\theta} = e^{jkh_1\cos\theta}\left(1 + \rho\, e^{-j\Delta}\right) \\ |A(\theta)|^2 &= \left|1 + \rho\, e^{-j\Delta}\right|^2, \quad \text{where} \quad \Delta = 2kh_1\cos\theta \end{aligned} \tag{19.3.6}$$

The gain of the transmitting antenna becomes $G_{\mathrm{tot}}(\theta) = |A(\theta)|^2 G(\theta)$, where $G(\theta)$ is the gain with the ground absent. For the common case of low grazing angles, or $\rho = -1$, the array factor becomes:

$$|A(\theta)|^2 = \left|1 - e^{-j\Delta}\right|^2 = 2 - 2\cos(\Delta) = 4\sin^2\left(\frac{\Delta}{2}\right)$$

At the location of the mobile antenna which is at height h_2, the geometry of the figure implies that $\cos\theta = h_2/r$. Thus, we have $\Delta = 2kh_1\cos\theta = 2kh_1h_2/r$, and

$$|A(\theta)|^2 = 4\sin^2\left(\frac{\Delta}{2}\right) \simeq \Delta^2 = \left(\frac{2kh_1h_2}{r}\right)^2$$

where we assumed that $kh_1h_2/r \ll 1$ and used the approximation $\sin x \simeq x$. Therefore, for fixed antenna heights h_1, h_2, the gain at the location of the receiving antenna drops like $1/r^2$. This is in addition to the $1/r^2$ drop arising from the power density. Thus, the presence of the ground reflection causes the overall power density at the receiving antenna to drop like $1/r^4$ instead of $1/r^2$.

For two antennas pointing towards the maximum gain of each other, the Friis transmission formula must be modified to read:

$$\frac{P_2}{P_1} = G_1G_2\left(\frac{\lambda}{4\pi r}\right)^2 \left|1 + \rho\, e^{-j\Delta}\right|^2, \qquad \Delta = \frac{2kh_1h_2}{r} = \frac{4\pi h_1h_2}{\lambda r} \tag{19.3.7}$$

The direct and ground-reflected rays are referred to as the *space wave*. When both antennas are close to the ground, one must also include a term in $A(\theta)$ due to the so-called *Norton surface wave* [1201–1206]:

$$A(\theta) = \underbrace{1 + \rho\, e^{-j\Delta}}_{\text{space wave}} + \underbrace{(1-\rho)Fe^{-j\Delta}}_{\text{surface wave}}$$

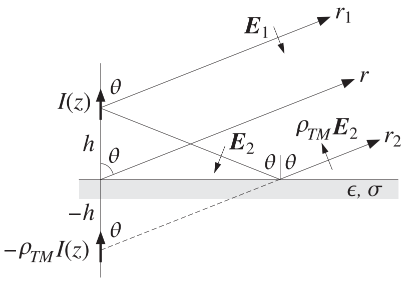
where F is an attenuation coefficient that, for $kr \gg 1$, can be approximated by [1193]:

$$F = \frac{\sin^2\alpha}{jkr(\cos\alpha + u)^2}\,, \qquad u = \frac{1}{n^2}\sqrt{n^2 - \sin^2\alpha}$$

At grazing angles, the space-wave terms of $A(\theta)$ tend to cancel and the surface wave becomes the only means of propagation. A historical review of the ground-wave propagation problem and some of its controversies can be found in [1191]. □

Example 19.3.6: *Vertical Dipole Antenna over Imperfect Ground.* Consider a vertical linear antenna at a height h over ground as shown below. When the observation point is far from the antenna, the direct and reflected rays r_1 and r_2 will be almost parallel to each other, forming an angle θ with the vertical. The incidence angle α of the previous example is then $\alpha = \theta$, so that the TM reflection coefficient is:

$$\rho_{TM} = \frac{\sqrt{n^2 - \sin^2\theta} - n^2\cos\theta}{\sqrt{n^2 - \sin^2\theta} + n^2\cos\theta}\,, \qquad n^2 = \epsilon_r - j\frac{\eta_0}{2\pi}\sigma\lambda$$



The relative permittivity $\epsilon_r = \epsilon/\epsilon_0$ and conductivity σ (in units of S/m) are given below for some typical grounds and typical frequencies:[†]

[†] ITU Recommendation ITU-R P.527-3 on the "Electrical Characteristics of the Surface of the Earth," 1992.

	1 MHz		100 MHz		1 GHz	
ground type	ϵ_r	σ	ϵ_r	σ	ϵ_r	σ
very dry ground	3	10^{-4}	3	10^{-4}	3	1.5×10^{-4}
medium dry ground	15	10^{-3}	15	1.5×10^{-3}	15	3.5×10^{-3}
wet ground	30	10^{-2}	30	1.5×10^{-2}	30	1.5×10^{-1}
fresh water	80	3×10^{-3}	80	5×10^{-3}	80	1.5×10^{-1}
sea water	70	5	70	5	70	5

According to Eq. (16.1.6), the electric fields $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ along the direct and reflected rays will point in the direction of their respective polar unit vector $\hat{\boldsymbol{\theta}}$, as seen in the above figure. According to the sign conventions of Sec. 7.2, the reflected field $\rho_{TM}\boldsymbol{E}_2$ will be pointing in the $-\hat{\boldsymbol{\theta}}$ direction, opposing $\boldsymbol{E}_1$. The net field at the observation point will be:

$$\boldsymbol{E}=\boldsymbol{E}_1-\rho_{TM}\boldsymbol{E}_2=\hat{\boldsymbol{\theta}}\,jk\eta\,\frac{e^{-jkr_1}}{4\pi r_1}F_z(\theta)\sin\theta-\hat{\boldsymbol{\theta}}\,jk\eta\,\frac{e^{-jkr_2}}{4\pi r_2}\rho_{TM}F_z(\theta)\sin\theta$$

where $\boldsymbol{F}(\theta)=\hat{\mathbf{z}}F_z(\theta)$ is the assumed radiation vector of the linear antenna. Thus, the reflected ray appears to have originated from an image current $-\rho_{TM}I(z)$. Using the approximations $r_1=r-h\cos\theta$ and $r_2=r+h\cos\theta$ in the propagation phase factors e^{-jkr_1} and e^{-jkr_2}, we obtain for the net electric field at the observation point (r,θ):

$$\boldsymbol{E}=\hat{\boldsymbol{\theta}}\,jk\eta\,\frac{e^{-jkr}}{4\pi r}F_z(\theta)\sin\theta\left[e^{jkh\cos\theta}-\rho_{TM}e^{-jkh\cos\theta}\right]$$

It follows that the (unnormalized) gain will be:

$$g(\theta)=\left|F_z(\theta)\sin\theta\right|^2\left|1-\rho_{TM}(\theta)e^{-2jkh\cos\theta}\right|^2$$

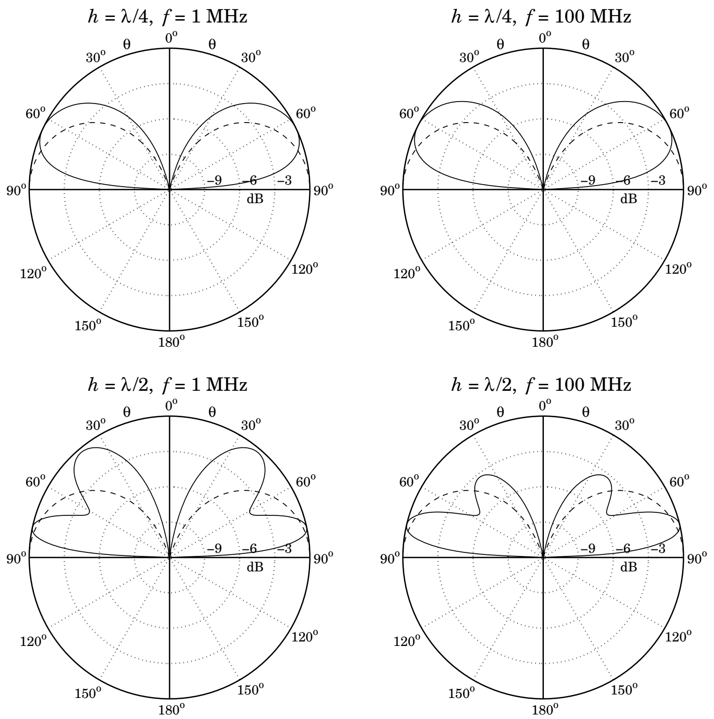
The results of the previous example are obtained if we set $\rho=-\rho_{TM}$. For a Hertzian dipole, we may replace $F_z(\theta)$ by unity. For a half-wave dipole, we have:

$$g(\theta)=\left|\frac{\cos(0.5\pi\cos\theta)}{\sin\theta}\right|^2\left|1-\rho_{TM}(\theta)e^{-2jkh\cos\theta}\right|^2$$

Fig. 19.3.5 shows the resulting gains for a half-wave dipole at heights $h=\lambda/4$ and $h=\lambda/2$ and at frequencies $f=1$ MHz and $f=100$ MHz. The ground parameters correspond to the *medium dry* case of the above table. The dashed curves represent the gain of a single dipole, that is, $G(\theta)=\left|\cos(0.5\pi\cos\theta)/\sin\theta\right|^2$.

The following MATLAB code illustrates the generation of these graphs:

```
sigma=1e-3; ep0=8.854e-12; er=15; f=1e6; h = 1/4;
n2 = er - j*sigma/ep0/2/pi/f;

th = linspace(0,pi/2,301); c =cos(th); s2 = sin(th).^2;
rho = (sqrt(n2-s2) - n2*c)./(sqrt(n2-s2) + n2*c);
A = 1 - rho .* exp(-j*4*pi*h*cos(th));               % array factor
G = cos(pi*cos(th)/2)./sin(th); G(1)=0;              % half-wave dipole gain

g = abs(G.*A).^2;  g = g/max(g);                     % normalized gain

dbp(th, g, 30, 12);                                  % polar plot in dB
```


$h=\lambda/4,\ f=1$ MHz

$h=\lambda/4,\ f=100$ MHz

$h=\lambda/2,\ f=1$ MHz

$h=\lambda/2,\ f=100$ MHz

Fig. 19.3.5 Vertical dipole over imperfect ground

Thus, the presence of the ground significantly alters the angular gain of the dipole. For the case $h=\lambda/2$, we observe the presence of grating lobes, arising because the effective separation between the dipole and its image is $2h>\lambda/2$.

The number of grating lobes increases with the height h. These can be observed by running the above example code with $f=1$ GHz (i.e., $\lambda=30$ cm) for a cell phone held vertically at a height of $h=6\lambda=1.8$ meters. ☐

19.4 One-Dimensional Arrays

Next, we consider uniformly-spaced one-dimensional arrays. An array along the x-axis (see Fig. 19.3.4) with elements positioned at locations x_n, $n=0,1,2,\dots$, will have displacement vectors $\mathbf{d}_n=x_n\hat{\mathbf{x}}$ and array factor:

$$A(\theta,\phi)=\sum_n a_n e^{j\mathbf{k}\cdot\mathbf{d}_n}=\sum_n a_n e^{jk_x x_n}=\sum_n a_n e^{jkx_n\sin\theta\cos\phi}$$

where we set $k_x=k\sin\theta\cos\phi$. For equally-spaced arrays, the element locations are $x_n=nd$, where d is the distance between elements. In this case, the array factor becomes:

$$\boxed{A(\theta,\phi)=\sum_n a_n e^{jnkd\sin\theta\cos\phi}} \qquad (19.4.1)$$

Because the angular dependence comes through the factor $k_x d = kd\sin\theta\cos\phi$, we are led to define the variable:

$$\boxed{\psi = k_x d = kd\sin\theta\cos\phi} \qquad \text{(digital wavenumber)} \qquad (19.4.2)$$

Then, the array factor may be thought of as a function of ψ:

$$\boxed{A(\psi)=\sum_n a_n e^{j\psi n}} \qquad \text{(array factor in digital wavenumber space)} \qquad (19.4.3)$$

The variable ψ is a normalized version of the wavenumber k_x and is measured in units of radians per (space) sample. It may be called a normalized *digital wavenumber*, in analogy with the time-domain normalized *digital frequency* $\omega = \Omega T = 2\pi f/f_s$, which is in units of radians per (time) sample.[†] The array factor $A(\psi)$ is the wavenumber version of the frequency response of a digital filter defined by

$$A(\omega)=\sum_n a_n e^{-j\omega n} \qquad (19.4.4)$$

We note the difference in the sign of the exponent in the definitions (19.4.3) and (19.4.4). This arises from the difference in defining time-domain and space-domain Fourier transforms, or from the difference in the sign for a plane wave, that is,

$$e^{j\omega t - j\mathbf{k}\cdot\mathbf{r}}$$

The wavenumber ψ is defined similarly for arrays along the y- or z-directions. In summary, we have the definitions:

$$\begin{aligned}
\psi &= k_x d = kd\sin\theta\cos\phi && \text{(array along } x\text{-axis)}\\
\psi &= k_y d = kd\sin\theta\sin\phi && \text{(array along } y\text{-axis)}\\
\psi &= k_z d = kd\cos\theta && \text{(array along } z\text{-axis)}
\end{aligned} \qquad (19.4.5)$$

The array factors for the y- and z-axis arrays shown in Fig. 19.1.1 will be:

$$\begin{aligned}
A(\theta,\phi) &= \sum_n a_n e^{jk_y y_n} = \sum_n a_n e^{jky_n\sin\theta\sin\phi}\\
A(\theta,\phi) &= \sum_n a_n e^{jk_z z_n} = \sum_n a_n e^{jkz_n\cos\theta}
\end{aligned}$$

where $y_n = nd$ and $z_n = nd$. More generally, for an array along some arbitrary direction, we have $\psi = kd\cos\gamma$, where γ is the angle measured from the direction of the array. The two most commonly used conventions are to assume either an array along the z-axis, or an array along the x-axis and measure its array factor only on the xy-plane, that is, at polar angle $\theta = 90^{\circ}$. In these cases, we have:

$$\begin{aligned}
\psi &= k_x d = kd\cos\phi && \text{(array along } x\text{-axis, with } \theta = 90^{\circ})\\
\psi &= k_z d = kd\cos\theta && \text{(array along } z\text{-axis)}
\end{aligned} \qquad (19.4.6)$$

[†] Here, Ω denotes the physical frequency in radians/sec.

For the x-array, the azimuthal angle varies over $-\pi \le \phi \le \pi$, but the array response is symmetric in ϕ and can be evaluated only for $0 \le \phi \le \pi$. For the z-array, the polar angle varies over $0 \le \theta \le \pi$.

In analogy with time-domain DSP, we may also define the spatial analog of the z-plane by defining the variable $z = e^{j\psi}$ and the corresponding z-transform:

$$\boxed{A(z)=\sum_n a_n z^n} \qquad \text{(array factor in spatial } z\text{-domain)} \qquad (19.4.7)$$

The difference in sign between the space-domain and time-domain definitions is also evident here, where the expansion is in powers of z^n instead of z^{-n}. The array factor $A(\psi)$ may be called the *discrete-space Fourier transform* (DSFT) of the array weighting sequence a_n, just like the discrete-time Fourier transform (DTFT) of the time-domain case. The corresponding inverse DSFT is obtained by

$$\boxed{a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\psi)\,e^{-j\psi n}d\psi} \qquad \text{(inverse DSFT)} \qquad (19.4.8)$$

This inverse transform forms the basis of most design methods for the array coefficients. As we mentioned earlier, such methods are identical to the methods of designing FIR filters in DSP. Various correspondences between the fields of array processing and time-domain digital signal processing are shown in Table 19.4.1.

Example 19.4.1: The array factors and z-transforms for Example 19.3.1 are for the three choices for the coefficients:

$$\begin{aligned}
A(\psi) &= 1 + e^{j\psi}\,, & A(z) &= 1 + z\\
A(\psi) &= 1 - e^{j\psi}\,, & A(z) &= 1 - z\\
A(\psi) &= 1 - je^{j\psi}\,, & A(z) &= 1 - jz
\end{aligned}$$

where $z = e^{j\psi}$ and $\psi = kd\cos\phi$. □

19.5 Visible Region

Because the correspondence from the physical angle-domain to the wavenumber ψ-domain is through the mapping (19.4.5) or (19.4.6), there are some additional subtleties that arise in the array processing case that do not arise in time-domain DSP. We note first that the array factor $A(\psi)$ is periodic in ψ with period 2π, and therefore, it is enough to know it within one Nyquist interval, that is, $-\pi \le \psi \le \pi$.

However, the actual range of variation of ψ depends on the value of the quantity $kd = 2\pi d/\lambda$. As the azimuthal angle ϕ varies from 0° to 180°, the quantity $\psi = kd\cos\phi$, defined in Eq. (19.4.6), varies from $\psi = kd$ to $\psi = -kd$. Thus, the overall range of variation of ψ—called the *visible region*—will be:

$$-kd \le \psi \le kd \qquad \text{(visible region)} \qquad (19.5.1)$$

discrete-time signal processing	discrete-space array processing
time-domain sampling $t_n = nT$	space-domain sampling $x_n = nd$
sampling time interval T	sampling space interval d
sampling rate $1/T$ [samples/sec]	sampling rate $1/d$ [samples/meter]
frequency Ω	wavenumber k_x
digital frequency $\omega = \Omega T$	digital wavenumber $\psi = k_x d$
Nyquist interval $-\pi \le \omega \le \pi$	Nyquist interval $-\pi \le \psi \le \pi$
sampling theorem $\Omega \le \pi/T$	sampling theorem $k_x \le \pi/d$
spectral images	grating lobes or fringes
frequency response $A(\omega)$	array factor $A(\psi)$
z-domain $z = e^{j\omega}$	z-domain $z = e^{j\psi}$
transfer function $A(z)$	transfer function $A(z)$
DTFT and inverse DTFT	DSFT and inverse DSFT
pure sinusoid $e^{j\omega_0 n}$	narrow beam $e^{-j\psi_0 n}$
windowed sinusoid $w(n)e^{j\omega_0 n}$	windowed narrow beam $w(n)e^{-j\psi_0 n}$
resolution of multiple sinusoids	resolution of multiple beams
frequency shifting by AM modulation	phased array scanning
filter design by window method	array design by window method
bandpass FIR filter design	angular sector array design
frequency-sampling design	Woodward-Lawson design
DFT	Blass matrix
FFT	Butler matrix

Table 19.4.1 Duality between time-domain and space-domain signal processing.

The total width of this region is $\psi_{\text{vis}} = 2kd$. Depending on the value of kd, the visible region can be less, equal, or more than one Nyquist interval:

$$\begin{array}{lllll} d < \lambda/2 & \Rightarrow & kd < \pi & \Rightarrow & \psi_{\text{vis}} < 2\pi \quad \text{(less than Nyquist)} \\ d = \lambda/2 & \Rightarrow & kd = \pi & \Rightarrow & \psi_{\text{vis}} = 2\pi \quad \text{(full Nyquist)} \\ d > \lambda/2 & \Rightarrow & kd > \pi & \Rightarrow & \psi_{\text{vis}} > 2\pi \quad \text{(more than Nyquist)} \end{array} \tag{19.5.2}$$

The visible region can also be viewed as that part of the unit circle covered by the angle range (19.5.1), as shown in Fig. 19.5.1. If $kd < \pi$, the visible region is the arc $z_a z z_b$ with the point $z = e^{j\psi}$ moving *clockwise* from z_a to z_b as ϕ varies from 0 to π. In the case $kd = \pi$, the starting and ending points, z_a and z_b, coincide with the $\psi = \pi$ point on the circle and the visible region becomes the entire circle. If $kd > \pi$, the visible region is one complete circle starting and ending at z_a and then continuing on to z_b.

In all cases, the inverse transform (19.4.8) requires that we know $A(\psi)$ over one complete Nyquist interval. Therefore, in the case $kd < \pi$, we must specify appropriate values of the array factor $A(\psi)$ over the invisible region.

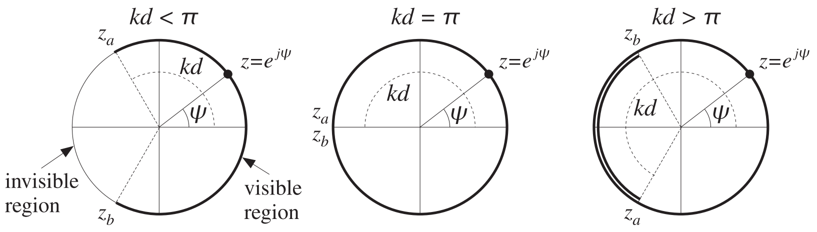


Fig. 19.5.1 Visible regions on the unit circle.

19.6 Grating Lobes

In the case $kd > \pi$, the values of $A(\psi)$ are over-specified and repeat over the visible region. This can give rise to *grating lobes* or *fringes*, which are mainbeam lobes in directions other than the desired one. We saw some examples in Figs. 19.3.1 and 19.3.2.

Grating lobes are essentially the *spectral images* generated by the sampling process (in this case, sampling in space.) In ψ-space, these images fall in Nyquist intervals other than the central one.

The number of grating lobes in an array pattern is the number of complete Nyquist intervals fitting within the width of the visible region, that is, $m = \psi_{\text{vis}}/2\pi = kd/\pi = 2d/\lambda$. For example in Fig. 19.3.2, the number of grating lobes are $m = 4, 8, 16$ for $d = 2\lambda, 4\lambda, 8\lambda$ (the two endfire lobes count as one.)

In most array applications grating lobes are undesirable and can be avoided by requiring that $kd < 2\pi$, or $d < \lambda$. It should be noted, however, that this condition does not necessarily avoid aliasing—it only avoids grating lobes. Indeed, if d is in the range $\lambda/2 < d < \lambda$, or, $\pi < kd < 2\pi$, part of the Nyquist interval repeats as shown in Fig. 19.5.1. To completely avoid repetitions, we must have $d \le \lambda/2$, which is equivalent to the *sampling theorem* condition $1/d \ge 2/\lambda$.

Grating lobes are desirable and useful in interferometry applications, such as *radio interferometry* used in radio astronomy. A simple interferometer is shown in Fig. 19.6.1. It consists of an array of two antennas separated by $d \gg \lambda$, so that hundreds or even thousands of grating lobes appear.

These lobes are extremely narrow allowing very small angular resolution of radio sources in the sky. The receiver is either an adder or a cross-correlator of the two antenna outputs. For an adder and identical antennas with equal weights, the output will be proportional to the array gain:

$$g(\phi) = \left|1 + e^{jkd\cos\phi}\right|^2 = 2 + 2\cos(kd\cos\phi)$$

For a cross-correlator, the output will be proportional to $\cos(\Omega\tau)$, where τ is the time delay between the received signals. This delay is the time it takes the wavefront to travel the distance $d\cos\phi$, as shown in Fig. 19.6.1, that is, $\tau = (d\cos\phi)/c$. Therefore,

$$\cos(\Omega\tau) = \cos\left(\frac{2\pi f d\cos\phi}{c}\right) = \cos(kd\cos\phi)$$

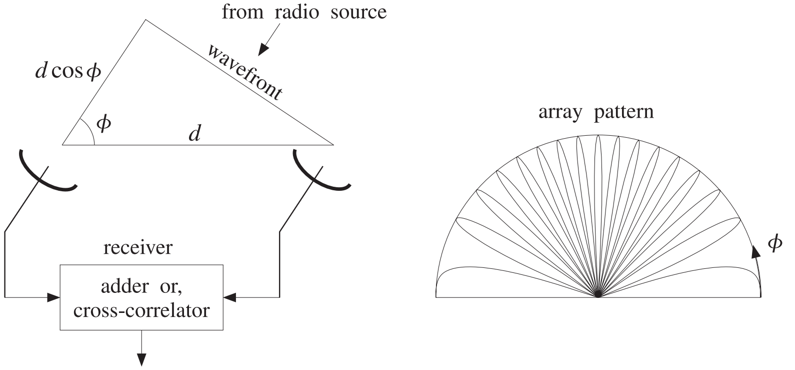


Fig. 19.6.1 Two-element interferometer and typical angular pattern.

In either case, the output is essentially $\cos(kd\cos\phi)$, and thus, exhibits the grating-lobe behavior. Cross-correlating interferometers are more widely used because they are more broadband.

The Very Large Array (VLA) radio telescope in New Mexico consists of 27 dish antennas with 25-m diameters. The antennas are on rails extending in three different directions to distances of up to 21 km. For each configuration, the number of possible interferometer pairs of antennas is $27(27-1)/2=351$. These 351 outputs can be used to make a "radio" picture of the source. The achievable resolution is comparable to that of optical telescopes (about 1 arc second.)

The Very Long Baseline Array (VLBA) consists of ten 25-m antennas located throughout the continental US, Puerto Rico, and Hawaii. The antennas are not physically connected to each other. Rather, the received signals at each antenna are digitally recorded, with the antennas being synchronized with atomic frequency standards, and then the recorded signals are digitally cross-correlated and processed off-line. The achievable resolution is about one milli-arc-second.

We note finally that in an interferometer, the angular pattern of each antenna element must also be taken into account because it multiplies the array pattern.

Example 19.6.1: In Fig. 19.3.2, we assumed isotropic antennas. Here, we look at the effect of the element patterns. Consider an array of two identical z-directed half-wavelength dipole antennas positioned along the z-axis at locations $z_0=0$ and $z_1=d$. The total polar gain pattern will be the product of the array gain factor and the gain of each dipole:

$$g_{\text{tot}}(\theta)=|A(\theta)|^2 g_{\text{dipole}}(\theta)=\big|a_0+a_1e^{jkd\cos\theta}\big|^2\left|\frac{\cos(0.5\pi\cos\theta)}{\sin\theta}\right|^2$$

Fig. 19.6.2 shows the effect of the element pattern for the case $d=8\lambda$ and uniform weights $\mathbf{a}=[a_0,a_1]=[1,1]$. The figure on the left represents the array factor, with the element pattern superimposed (dashed gain). On the right is the total gain.

The MATLAB code used to generate the right graph was as follows:

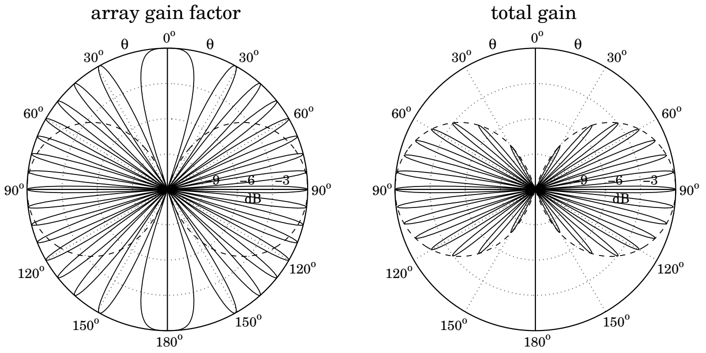


Fig. 19.6.2 Grating lobes of two half-wavelength dipoles separated by $d=8\lambda$.

```
d=8; a=[1,1];
[g, th] = gain1d(d, a, 400);
gdip = dipole(0.5, 400);
gtot = g .* gdip;
dbp(th, gtot, 30, 12);
dbadd(1, '--', th, gdip, 30, 12);
```

□

19.7 Uniform Arrays

The simplest one-dimensional array is the uniform array having equal weights. For an array of N isotropic elements at locations $x_n=nd$, $n=0,1,\dots,N-1$, we define:

$$\mathbf{a}=[a_0,a_1,\dots,a_{N-1}]=\frac{1}{N}[1,1,\dots,1] \tag{19.7.1}$$

so that the sum of the weights is unity. The corresponding array polynomial and array factor are:

$$\begin{aligned} A(z)&=\frac{1}{N}\big[1+z+z^2+\cdots+z^{N-1}\big]=\frac{1}{N}\,\frac{z^N-1}{z-1}\\ A(\psi)&=\frac{1}{N}\big[1+e^{j\psi}+e^{2j\psi}+\cdots+e^{(N-1)j\psi}\big]=\frac{1}{N}\,\frac{e^{jN\psi}-1}{e^{j\psi}-1}\end{aligned} \tag{19.7.2}$$

where $z=e^{j\psi}$ and $\psi=kd\cos\phi$ for an array along the x-axis and look direction on the xy-plane. We may also write $A(\psi)$ in the form:

$$\boxed{A(\psi)=\frac{\sin\left(\dfrac{N\psi}{2}\right)}{N\sin\left(\dfrac{\psi}{2}\right)}\,e^{j(N-1)\psi/2}} \qquad \text{(uniform array)} \tag{19.7.3}$$

The array factor (19.7.2) is the spatial analog of a lowpass FIR averaging filter in discrete-time DSP. It may also be viewed as a window-based narrow-beam design using a *rectangular* window. From this point of view, Eq. (19.7.3) is the DSFT of the rectangular window.

The array factor has been normalized to have unity gain at dc, that is, at zero wavenumber $\psi = 0$, or at the broadside azimuthal angle $\phi = 90^{\text{o}}$. The normalized power gain of the array will be:

$$g(\phi) = |A(\psi)|^2 = \left|\frac{\sin(N\psi/2)}{N\sin(\psi/2)}\right|^2 = \left|\frac{\sin\big((Nkd/2)\cos\phi\big)}{N\sin\big((kd/2)\cos\phi\big)}\right|^2 \tag{19.7.4}$$

Although (19.7.2) defines the array factor for all ψ over one Nyquist interval, the actual visible region depends on the value of kd.

Fig. 19.7.1 shows $A(\psi)$ evaluated only over its visible region for an 8-element ($N = 8$) array, for the following three choices of the element spacing: $d = 0.25\lambda$, $d = 0.5\lambda$, and $d = \lambda$. The following MATLAB code generates the last two graphs:

```
d=1; N=8;
a = uniform(d, 90, N);
[g, phi] = gain1d(d, a, 400);
A = sqrt(g);
psi = 2*pi*d*cos(phi);
plot(psi/pi, A);
figure(2);
dbz(phi, g, 45, 20);
```

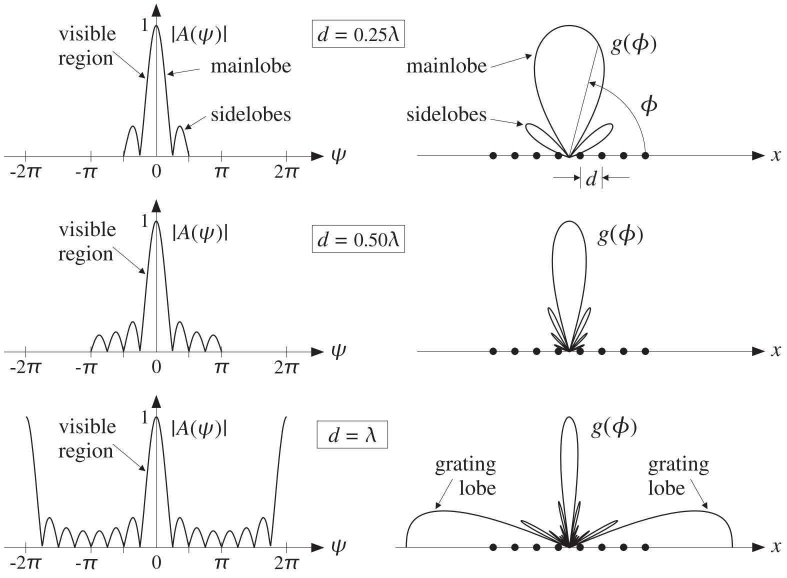


Fig. 19.7.1 Array factor and angular pattern of 8-element uniform array.

As ϕ varies from 0^{o} to 180^{o}, the visible regions for the three cases are:

$$\begin{array}{lll} d = 0.25\lambda, & \psi = (\pi/2)\cos\phi & \Rightarrow \quad -\pi/2 \le \psi \le \pi/2 \\ d = 0.5\lambda, & \psi = \pi\cos\phi & \Rightarrow \quad -\pi \le \psi \le \pi \\ d = \lambda, & \psi = 2\pi\cos\phi & \Rightarrow \quad -2\pi \le \psi \le 2\pi \end{array}$$

Thus, in the first case the visible region is only half of the Nyquist interval; in the second case, it is the full interval; and in the third case, the Nyquist interval is covered twice, and therefore, grating lobes will appear. Because $\psi = 2\pi\cos\phi$, the grating lobes at $\psi = \pm 2\pi$ correspond to the endfire angles of $\phi = 0^{\text{o}}$ and 180^{o} (the larger width of the endfire lobes is explained in Sec. 19.10.)

The $N - 1$ zeros of the array polynomial $A(z)$ are the N-th roots of unity, except for the root at $z = 1$, that is,

$$z_i = e^{j\psi_i}, \quad \psi_i = \frac{2\pi i}{N}, \quad i = 1, 2, \dots, N-1$$

Because these zeros lie on the unit circle, they will correspond to nulls in the angular pattern, as long as they lie in the visible region. For $d = 0.25\lambda$, and in general for any $d < \lambda/2$, only a subset of these zeros will fall in the visible region. The zeros of the 8-element array patterns of Fig. 19.7.1 are shown in Fig. 19.7.2.

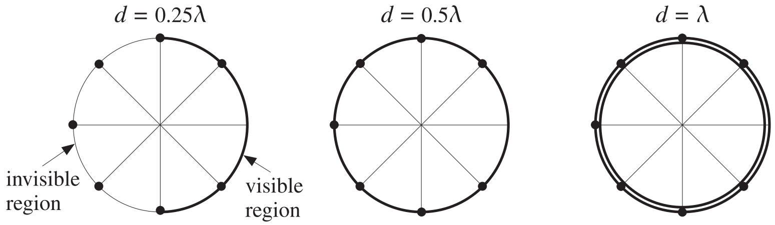


Fig. 19.7.2 Zero locations and visible regions of 8-element uniform array.

The two most important features of the uniform array are its 3-dB beamwidth $\Delta\psi_{\text{3dB}}$, or $\Delta\phi_{\text{3dB}}$ in angle-space, and its sidelobe level R. These parameters are shown in Fig. 19.7.3, for an 8-element uniform array with $d = 0.5\lambda$.

For N larger than about 5–6, the sidelobe level becomes independent of N and has the limiting value of $R = 13$ dB. Similarly, the beamwidth in ψ-space—defined as the full width of the mainlobe at the half-power level—takes the simple form:

$$\boxed{\Delta\psi_{\text{3dB}} = 0.886\,\frac{2\pi}{N}} \qquad \text{(3-dB width in } \psi\text{-space)} \tag{19.7.5}$$

The first nulls in the array factor about the mainlobe are at $\pm\psi_1 = \pm 2\pi/N$, and therefore, $2\pi/N$ represents half of the base of the mainlobe.

The 3-dB width $\Delta\phi_{\text{3dB}}$ in angle space can be obtained by differentiating the equation $\psi = kd\cos\phi$, that is, $d\psi = (\partial\psi/\partial\phi)d\phi = (-kd\sin\phi)d\phi$. Evaluating the derivative at broadside ($\phi = 90^{\text{o}}$) and assuming a narrow mainlobe, we have:

$$\Delta\psi_{\text{3dB}} = \left|\frac{\partial\psi}{\partial\phi}\right| \Delta\phi_{\text{3dB}} = kd\,\Delta\phi_{\text{3dB}}$$

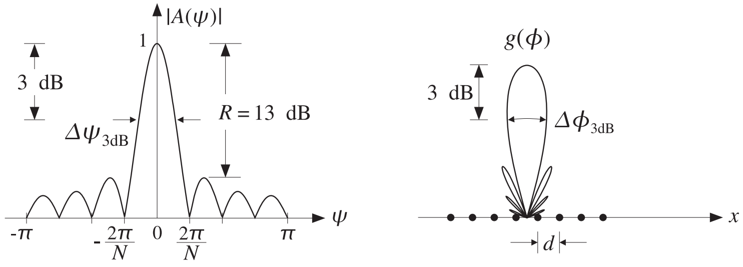


Fig. 19.7.3 Mainlobe width and sidelobe level of uniform array.

Solving for $\Delta\phi_{3\mathrm{dB}}$, we obtain $\Delta\phi_{3\mathrm{dB}} = \Delta\psi_{3\mathrm{dB}}/(kd) = 0.886(2\pi/N)/(2\pi d/\lambda)$, or

$$\boxed{\Delta\phi_{3\mathrm{dB}} = 0.886\,\frac{\lambda}{Nd}} \qquad \text{(3-dB width at broadside)} \qquad (19.7.6)$$

The mainlobe beamwidth gets narrower with increasing N, while the relative sidelobe level remains the same. To achieve better (lower) sidelobe levels, one must use non-uniform weights obtained from non-rectangular windows.

The quantity $D = Nd$ is the effective *aperture* of the array. Thus, we recognize Eq. (19.7.6) as the classical *Rayleigh limit* on the resolving power of an optical system, which states that the *angular resolution* achieved by an aperture of length D is essentially λ/D.

The beamwidth expression (19.7.5) and the 13-dB sidelobe level can be justified as follows. The peak of the first sidelobe occurs approximately half-way between the first two nulls, that is, at $\psi = 3\pi/N$. More precisely, it occurs at $\psi = 2.8606\pi/N$. Thus, the sidelobe level in dB will be:

$$\begin{aligned} R &= -20\log_{10}\left|\frac{A(\psi)}{A(0)}\right|_{\psi=2.8606\pi/N} = -20\log_{10}\left|\frac{\sin(1.4303\pi)}{N\sin(1.4303\pi/N)}\right| \\ &\simeq -20\log_{10}\left|\frac{\sin(1.4303\pi)}{N(1.4303\pi/N)}\right| = -20\log_{10}\left|\frac{\sin(1.4303\pi)}{1.4303\pi}\right| = 13.26\ \mathrm{dB} \end{aligned}$$

where we used the small-x approximation, $\sin x \simeq x$, in the denominator, which is justified when N is large. Setting $x = N\psi/2$, the sidelobe peak corresponds to the secondary maximum of the approximate array factor $\sin x/x$, which by differentiation leads to the equation $x = \tan x$, having solution $x = 1.4303\pi$, or $\psi = 2x/N = 2.8606\pi/N$.

The 3-dB width $\Delta\psi_{3\mathrm{dB}}$ is twice the 3-dB or half-power frequency ψ_3, defined to be the solution of the equation:

$$|A(\psi_3)|^2 = \left|\frac{\sin(N\psi_3/2)}{N\sin(\psi_3/2)}\right|^2 = \frac{1}{2}$$

Because ψ_3 is always smaller than $2\pi/N$, it will be small for large N, and therefore, we may make the same approximation in the denominator as above, giving the simplified equation:

$$\left|\frac{\sin(N\psi_3/2)}{N\psi_3/2}\right|^2 = \left|\frac{\sin x_3}{x_3}\right|^2 = \frac{1}{2}$$

where $x_3 = N\psi_3/2$. The quantity x_3 is determined to be the constant $x_3 = 0.443\pi$. Thus, $\psi_3 = 2x_3/N = 0.443(2\pi/N)$, and $\Delta\psi_{3\mathrm{dB}} = 2\psi_3 = 0.886(2\pi/N)$.

19.8 Array Directivity

The value of kd has an impact also on the directivity of an array. In the array processing literature, the directivity of an array is usually defined with reference to a z-directed array consisting of *isotropic* radiators. The wavenumber is $\psi = kd\cos\theta$ and the maximum of the array factor is assumed to occur at broadside $\theta = 90^{\circ}$, or $\psi = 0$. This basically means that the array factor will have a lowpass shape as a function of ψ, with a maximum value at dc given by

$$|A(0)| = \left|\sum_{n=0}^{N-1} a_n\right|$$

It follows that the normalized power gain of the array will be:

$$g(\theta) = c|A(\theta)|^2$$

where $c = 1/|A(0)|^2$. The corresponding beam solid angle will be:

$$\Delta\Omega = 2\pi\int_0^{\pi} g(\theta)\sin\theta\,d\theta = 2\pi\int_0^{\pi} c|A(\theta)|^2\sin\theta\,d\theta$$

Changing variables of integration from θ to ψ, which varies over the visible region (19.5.1), we obtain:

$$\Delta\Omega = \frac{2\pi}{kd}\int_{-kd}^{kd} c|A(\psi)|^2\,d\psi = \frac{2\pi c}{kd}\int_{-kd}^{kd}\sum_{n,m} a_n a_m^* e^{j(n-m)\psi}\,d\psi$$

Performing the integration, we get

$$\Delta\Omega = 4\pi c\sum_{n,m} a_n a_m^*\,\frac{\sin\big(kd(n-m)\big)}{kd(n-m)}$$

Therefore, the directivity of the array becomes:

$$D = \frac{4\pi}{\Delta\Omega} = \frac{\left|\sum_n a_n\right|^2}{\sum_{n,m} a_n a_m^*\,\dfrac{\sin\big(kd(n-m)\big)}{kd(n-m)}} \qquad (19.8.1)$$

In the particular case of half-wavelength spacing $d = \lambda/2$ or $kd = \pi$, the sinc function acts as a delta function $\delta(n-m)$, and the sum simplifies into:

$$D = \frac{\left|\sum_{n=0}^{N-1} a_n\right|^2}{\sum_{n=0}^{N-1}|a_n|^2} \qquad (19.8.2)$$

The maximum of this quantity is reached when all the coefficients are equal to each other. The common value may be adjusted so that their sum is unity, that is:

$$a_n = \frac{1}{N}, \qquad n = 0, 1, \dots, N-1$$

The maximized value of D becomes:

$$D_{\max} = N \tag{19.8.3}$$

Thus, the uniform array with half-wavelength spacing achieves *maximum directivity* equal to the *number* of array elements. This result is analogous to finding the optimum N-tap lowpass FIR filter that minimizes the noise reduction ratio, that is, the sum of the squares of its coefficients.

For arbitrary spacing d, it is shown in Problem 19.6 that the optimum array vector $\mathbf{a} = [a_0, a_1, \dots, a_{N-1}]^T$ that maximizes (19.8.1), and the corresponding maximum directivity, are given by:

$$\mathbf{a} = A^{-1}\mathbf{u}, \qquad D_{\max} = \mathbf{u}^T A^{-1}\mathbf{u} \tag{19.8.4}$$

where $\mathbf{u} = [1, 1, \dots, 1]^T$ is a vector of N ones and A is the so-called *prolate matrix* [1099] with matrix elements:

$$A_{nm} = \frac{\sin\big(kd(n-m)\big)}{kd(n-m)}, \quad 0 \le n, m \le N-1 \tag{19.8.5}$$

The coefficients $\mathbf{a}$ may be renormalized such that their sum is unity. When d is an integer multiple of $\lambda/2$, the prolate matrix reduces to the $N\times N$ identity matrix, resulting into (19.8.3).

19.9 Array Steering

An array is typically designed to have maximum directive gain at broadside, that is, at $\phi = 90^{\text{o}}$ (for an array along the x-axis.) The maximum of the array factor $A(\psi)$ corresponds to $\psi = kd\cos\phi = 0$, so that $|A|_{\max} = |A(0)|$.

We wish to "electronically" rotate, or steer, the array pattern towards some other direction, say ϕ_0, without physically rotating it. The corresponding wavenumber at the desired look-direction will be:

$$\boxed{\psi_0 = kd\cos\phi_0} \qquad \text{(steering phase)} \tag{19.9.1}$$

Such steering operation can be achieved by *wavenumber translation* in ψ-space, that is, replacing the broadside pattern $A(\psi)$ by the translated pattern $A(\psi - \psi_0)$. Thus, we define:

$$\boxed{A'(\psi) = A(\psi - \psi_0)} \qquad \text{(steered array factor)} \tag{19.9.2}$$

and the translated wavenumber variable,

$$\boxed{\psi' = \psi - \psi_0 = kd(\cos\phi - \cos\phi_0)} \qquad \text{(steered wavenumber)} \tag{19.9.3}$$

Then, $A'(\psi) = A(\psi')$. The maximum of $A'(\psi)$ will coincide with the maximum of $A(\psi')$, which occurs at $\psi' = 0$, or equivalently at $\psi = \psi_0$, or at angle $\phi = \phi_0$. Fig. 19.9.1 illustrates this wavenumber translation process and the corresponding rotation of the angular pattern, for an 11-element uniform array with $d = \lambda/2$, steered from broadside to $\phi_0 = 60^{\text{o}}$. The MATLAB code for the last two graphs was:

```
d=0.5; N=11; ph0=60;
a = uniform(d, ph0, N);                 % steered uniform weights
[g, phi] = gain1d(d, a, 400);           % calculate normalized gain g(φ)
psi = 2*pi*d*cos(phi);                  % φ to ψ transformation
figure; plot(psi/pi, sqrt(g));          % plot in ψ space
figure; dbz(phi, g, 30, 20);            % azimuthal gain plot in dB
```

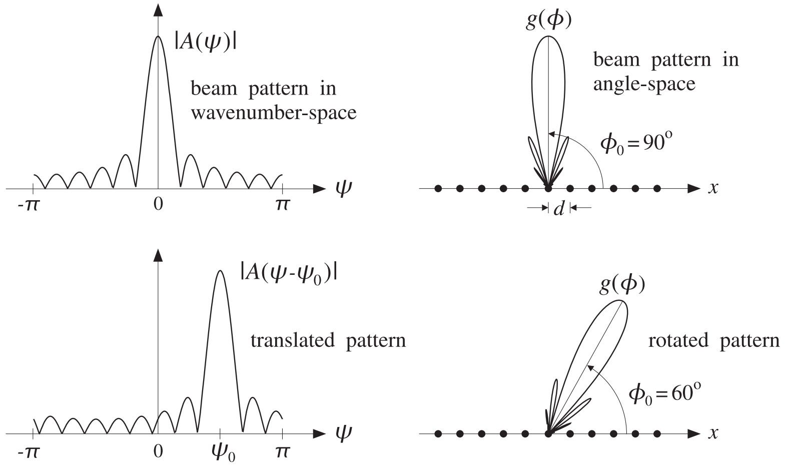


Fig. 19.9.1 Array steering or scanning by translation in wavenumber space.

It follows from the translation theorem of Fourier transforms that the weight coefficients a'_n of the translated pattern $A'(\psi)$ will be given by:

$$\boxed{a'_n = a_n e^{-j\psi_0 n}} \qquad \text{(steered array weights)} \tag{19.9.4}$$

so that we have:

$$A'(\psi) = \sum_n a'_n e^{j\psi n} = \sum_n a_n e^{j(\psi-\psi_0)n} = \sum_n a_n e^{j\psi' n} = A(\psi')$$

Because of the progressive phase factors $e^{-j\psi_0 n}$ in the weights a'_n, the steered or scanned array is sometimes called a *phased* or *scanning* array.

The time-domain version of array steering is AM modulation, in which a baseband signal is translated up in frequency by modulating with it a sinusoidal carrier, much like Eq. (19.9.4). Frequency translation is also used in DSP for mapping a lowpass filter into a bandpass one and for designing filter banks. We will use it in Sec. 20.4 to design arrays with angular sector patterns.

The MATLAB functions `steer.m` and `scan.m` of Appendix I can be used to implement Eq. (19.9.4). Their usage for even or odd number of array elements is discussed in Sec. 20.1.

Example 19.9.1: In Examples 19.3.1 and 19.3.2, we considered the three cases having progressive phases $\psi_0 = 0, \pi, \pi/2$. These may or may not correspond to a physical steering angle ϕ_0, depending on whether or not ψ_0 lies in the visible region.

In the case $\psi_0 = \pi$ and $d = 0.25\lambda$, we have $\psi = 0.5\pi\cos\phi$, and therefore it is not possible to find a solution for $0.5\pi\cos\phi_0 = \psi_0 = \pi$. However, the array factor does correspond to a pattern rotated towards endfire. This can be seen from the expression,

$$|A(\psi)| = |1 - e^{j\psi}| = 2\big|\sin(\psi/2)\big| = 2\big|\sin(0.25\pi\cos\phi)\big|$$

which is maximum towards endfire and minimum towards broadside. In the case $\psi_0 = \pi/2$ and $d = 0.25\lambda$, there is a solution to $0.5\pi\cos\phi_0 = \psi_0 = 0.5\pi$, that is, $\phi_0 = 0^{\circ}$, which corresponds to the maximum of the steered array.

In the case $\psi_0 = \pi$ and $d = 0.5\lambda$, we have $\psi = \pi\cos\phi$, and the solution to the equation $\pi\cos\phi_0 = \pi$ is $\phi_0 = 0^{\circ}$. However, because the phase $\psi_0 = \pi$ is indistinguishable from the phase $\psi_0 = -\pi$ (both lead to $e^{-j\psi_0} = -1$), we will also have the solution to $\pi\cos\phi_0 = -\pi$, which is $\phi_0 = 180^{\circ}$.

In the case $\psi_0 = \pi/2$ and $d = 0.5\lambda$, the solution to $\pi\cos\phi_0 = \pi/2$ is $\phi_0 = 60^{\circ}$, which corresponds to the maximum, as can be seen in Fig. 19.3.1.

In the case $\psi_0 = \pm\pi$ and $d = \lambda$, we have $\psi = 2\pi\cos\phi$, and the solutions to $2\pi\cos\phi_0 = \pm\pi$ are $\phi_0 = 60^{\circ}$ and 120°.

Finally, in the case $\psi_0 = \pi/2$ and $d = \lambda$, the solution to $2\pi\cos\phi_0 = \pi/2$ is $\phi^0 = 75.5^{\circ}$. However, there is another grating lobe maximum towards $\phi_0 = 138.6^{\circ}$, which corresponds to the solution of $2\pi\cos\phi_0 = -3\pi/2$. This is so because $\psi_0 = \pi/2$ and $\psi_0 = -3\pi/2$ are indistinguishable phases, both leading to $e^{-j\psi_0} = -j$. □

The concepts of visible region, beamwidth, and the condition for absence of grating lobes, translate with minor modifications to the case of a steered array. As the angle ϕ varies over $0^{\circ} \le \phi \le 180^{\circ}$, the translated wavenumber ψ' of Eq. (19.9.3) varies over the shifted visible region:

$$\boxed{-kd(1+\cos\phi_0) \le \psi' \le kd(1-\cos\phi_0)} \qquad \text{(shifted visible region)} \qquad (19.9.5)$$

where its total width is again $2kd$. The condition for absence of grating lobes is obtained with the help of the inequality:

$$|\psi'| \le kd|\cos\phi - \cos\phi_0| \le kd\big(|\cos\phi| + |\cos\phi_0|\big) \le kd\big(1 + |\cos\phi_0|\big)$$

To ensure no grating lobes, ψ' must remain strictly less than 2π, which results in the sufficient condition: $kd\big(1+|\cos\phi_0|\big) < 2\pi$, or replacing $kd = 2\pi d/\lambda$,

$$\boxed{d < \frac{\lambda}{1 + |\cos\phi_0|}} \qquad \text{(no grating lobes)} \qquad (19.9.6)$$

At broadside, $\phi_0 = 90^{\circ}$, this reduces to the earlier condition $d < \lambda$. At endfire, $\phi_0 = 0^{\circ}$ or 180°, it reduces to $d < \lambda/2$.

19.10 Array Beamwidth

Because the steered array has a mainlobe towards the direction ϕ_0, the beamwidth must be calculated by linearizing the map $\psi = kd\cos\phi$ about ϕ_0, that is,

$$\Delta\psi = \left|\frac{\partial\psi}{\partial\phi}\right|_{\phi_0} \Delta\phi = |-kd\sin\phi_0|\,\Delta\phi$$

which leads to the 3-dB beamwidth in angle-space:

$$\Delta\phi_{\text{3dB}} = \frac{1}{kd\sin\phi_0}\,\Delta\psi_{\text{3dB}}\,, \qquad \text{(3-dB width of steered array)} \qquad (19.10.1)$$

For window-based narrow-beam design methods, the beamwidth $\Delta\psi_{\text{3dB}}$ is approximately equal to the product of the beamwidth of the uniform array, Eq. (19.7.5), and a so-called *broadening factor* b, whose value depends on the choice of the window. Thus, we have:

$$\boxed{\Delta\psi_{\text{3dB}} = b\,\Delta\psi_{\text{3-dB, uniform}} = 0.886\,\frac{2\pi b}{N}} \qquad \text{(3-dB width in }\psi\text{-space)} \qquad (19.10.2)$$

Combining Eqs. (19.10.1) and (19.10.2) and replacing kd by $2\pi d/\lambda$, we get:

$$\Delta\phi_{\text{3dB}} = \frac{0.886}{\sin\phi_0}\,\frac{\lambda}{Nd}\,b\,, \qquad \text{(3-dB width in angle-space)} \qquad (19.10.3)$$

The 3-dB angles will be approximately $\phi_0 \pm \Delta\phi_{\text{3dB}}/2$. Because of the presence of $\sin\phi_0$ in the denominator, the beamwidth $\Delta\phi_{\text{3dB}}$ will broaden as the array is steered from broadside to endfire.

Exactly at endfire, $\phi_0 = 0^{\circ}$ or 180°, Eq. (19.10.3) fails and the beamwidth must be calculated by a different procedure. At $\phi_0 = 0^{\circ}$, the translated wavenumber $\psi' = \psi - \psi_0$ becomes $\psi' = kd(\cos\phi - 1)$. Using the approximation $\cos x = 1 - x^2/2$, we may relate the 3-dB angle ϕ_3 to the corresponding 3-dB wavenumber by:

$$\psi_3' = kd(\cos\phi_3 - 1) = kd\big((1 - \phi_3^2/2) - 1\big) = -\frac{1}{2}kd\phi_3^2$$

It follows that the 3-dB width in ψ-space will be $\Delta\psi_{\text{3dB}} = 2|\psi_3'| = kd\phi_3^2$. Solving for ϕ_3, we have $\phi_3 = \sqrt{\Delta\psi_{\text{3dB}}/kd}$. Thus, the 3-dB width in angle space will be $\Delta\phi_{\text{3dB}} = 2\phi_3$,

$$\Delta\phi_{\text{3dB}} = 2\sqrt{\frac{\Delta\psi_{\text{3dB}}}{kd}}\,, \qquad \text{(3-dB width at endfire)} \qquad (19.10.4)$$

The same expression also holds for endfire towards $\phi_0 = 180^{\circ}$. Replacing $\Delta\psi_{\text{3dB}}$ from Eq. (19.10.2), we find the width in angle space:

$$\Delta\phi_{\text{3dB}} = 2\sqrt{0.886\,\frac{\lambda}{Nd}\,b}\,, \qquad \text{(3-dB width in angle-space)} \qquad (19.10.5)$$

To summarize, the angular 3-dB width of the steered array can be computed in terms of the broadside 3-dB width in wavenumber space by:

$$\boxed{\Delta\phi_{3\text{dB}} = \begin{cases} \dfrac{1}{kd\sin\phi_0}\,\Delta\psi_{3\text{dB}}\,, & \text{for } 0^o < \phi_0 < 180^o \\[2ex] 2\sqrt{\dfrac{\Delta\psi_{3\text{dB}}}{kd}}\,, & \text{for } \phi_0 = 0^o,\ 180^o \end{cases}} \tag{19.10.6}$$

In particular, if Eq. (19.10.2) is used:

$$\boxed{\Delta\phi_{3\text{dB}} = \begin{cases} \dfrac{0.886}{\sin\phi_0}\,\dfrac{\lambda}{Nd}\,b\,, & \text{for } 0^o < \phi_0 < 180^o \\[2ex] 2\sqrt{0.886\,\dfrac{\lambda}{Nd}\,b}\,, & \text{for } \phi_0 = 0^o,\ 180^o \end{cases}} \tag{19.10.7}$$

In degrees, Eq. (19.10.7) reads as:

$$\Delta\phi_{3\text{dB}} = \begin{cases} \dfrac{50.76^o}{\sin\phi_0}\,\dfrac{\lambda}{Nd}\,b\,, & \text{for } 0^o < \phi_0 < 180^o \\[2ex] 107.86^o\sqrt{\dfrac{\lambda}{Nd}\,b}\,, & \text{for } \phi_0 = 0^o,\ 180^o \end{cases} \tag{19.10.8}$$

In some designs such as binomial arrays, it is easier to determine $\Delta\psi_{3\text{dB}}$ directly from the array factor $A(\psi)$. In other designs, it is more convenient to estimate $\Delta\psi_{3\text{dB}}$ using Eq. (19.10.2).

The broadening factor b depends on the choice of the window and its sidelobe level. The larger the sidelobe attenuation, the larger the broadening factor. Some examples of broadening factors for different windows are given as follows:

Rectangular:	$b = 1, \quad (R = 13\text{ dB})$
Hamming:	$b = 2, \quad (R = 40\text{ dB})$
Taylor-Kaiser [1090]:	$b = \dfrac{6(R+12)}{155}$
Dolph-Chebyshev [1088]:	$b = 1 + 0.636\left[\dfrac{2}{R_a}\cosh\left(\sqrt{\text{acosh}^2(R_a) - \pi^2}\right)\right]^2$

where R and R_a represent the sidelobe level in dB and absolute units, respectively,

$$\boxed{R = 20\log_{10}(R_a) \quad\Leftrightarrow\quad R_a = 10^{R/20}} \qquad \text{(sidelobe level)} \tag{19.10.9}$$

Here, R and R_a represent the *attenuation* of the sidelobe and, therefore, $R > 0$ and $R_a > 1$. The corresponding gain of the sidelobe relative to the mainlobe peak will be $R_a^{-1} = 10^{-R/20}$, which is less than one.

The MATLAB function `bwidth.m` of Appendix I implements Eq. (19.10.6). Its inputs are the quantities d, ϕ_0, $\Delta\psi_{3\text{dB}}$ and its output is the 3-dB width in degrees $\Delta\phi_{3\text{dB}}$. Its usage is:

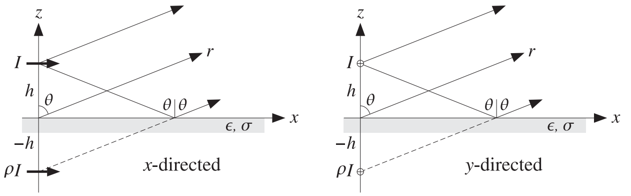
```
Dphi = bwidth(d, phi0, Dpsi);        % map Δψ beamwidth to Δφ beamwidth
```

19.11 Problems

19.1 Show that the modified Friis formula (19.3.7) for two antennas over imperfect ground takes the following frequency-independent form in the limit of low grazing angles and $h_1 h_2 \ll \lambda r$:

$$\frac{P_2}{P_1} = G_1 G_2\left(\frac{h_1 h_2}{r^2}\right)^2$$

19.2 Consider two horizontal dipoles I over imperfect ground, oriented along the x and y directions, as shown below. Show that the effect of the direct and ground-reflected rays can be obtained by considering an image dipole ρI.

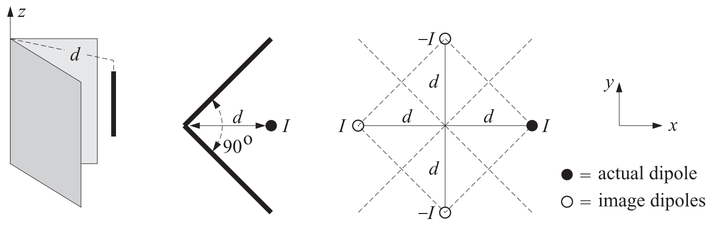


By considering the relative directions of the electric field along the direct and reflected rays, show that the resulting in array factor has the form:

$$A(\theta) = e^{jkh\cos\theta} + \rho\, e^{-jkh\cos\theta}$$

with $\rho = \rho_{TM}$ for the x-directed case and $\rho = \rho_{TE}$ for the y-directed one, where ρ_{TM}, ρ_{TE} are given by Eq. (7.4.4) with $n^2 = \epsilon_r - j60\sigma\lambda$.

19.3 A z-directed half-wave dipole is positioned in front of a 90^o corner reflector at a distance d from the corner, as shown below. The reflecting conducting sheets can be removed and replaced by three image dipoles of alternating signs, as shown.



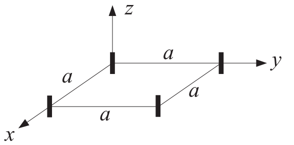
a. Thinking of the equivalent image problem as an array, determine an analytical expression for the array factor $A(\theta,\phi)$ as a function of the polar and azimuthal angles θ,ϕ.

b. For the values $d = 0.5\lambda$, $d = \lambda$, and $d = 1.5\lambda$, plot the azimuthal pattern $A(90^o,\phi)$ at polar angle $\theta = 90^o$ and for $-45^o \le \phi \le 45^o$.

c. For the cases $d = 0.5\lambda$ and $d = 1.5\lambda$, calculate the directivity D (in dB and in absolute units) and compare it with the directivity of a single half-wave dipole in the absence of the reflector.

d. Suppose that the corner reflector is flattened into a conducting sheet lying on the yz plane, i.e., the 90^o angle between the sheets is replaced by a 180^o angle. Repeat parts (a–c) in this case.

19.4 Four identical isotropic antennas are positioned on the xy-plane at the four corners of a square of sides a, as shown below. Determine the array factor $A(\phi)$ of this arrangement as a function of the azimuthal angle ϕ. (Assume the look direction is on the xy-plane.)



19.5 The array factor of a two-element array is given by:

$$g(\phi)=\left|a_0+a_1e^{j\psi}\right|^2=\frac{1+\sin\psi}{2}\,,\quad \psi=\frac{\pi}{2}\cos\phi$$

where ϕ is the azimuthal angle (assume $\theta = 90^o$) and ψ, the digital wavenumber. The array elements are along the x-axis at locations $x_0 = 0$ and $x_1 = d$.

a. What is the spacing d in units of λ? Determine the values of the array weights, $\mathbf{a} = [a_0, a_1]$, assuming that a_0 is real-valued and positive.

b. Determine the visible region and display it on the unit circle. Plot $|A(\psi)|^2$ versus ψ over the visible region. Based on this plot, make a rough sketch of the radiation pattern of the array (i.e., the polar plot of $g(\phi)$ versus $0 \le \phi \le 2\pi$).

c. Determine the exact 3-dB width of this array in angle space.

19.6 Defining the array vector $\mathbf{a}$ and the prolate matrix A via Eqs. (19.8.4) and (19.8.5), show that the directivity defined in Eq. (19.8.1) can be written in the compact form, where the dagger † indicates the conjugate transposed operation:

$$D=\frac{\left|\mathbf{u}^\dagger\mathbf{a}\right|^2}{\mathbf{a}^\dagger A\mathbf{a}} \tag{19.11.1}$$

a. Show that the maximum of D is attained for $\mathbf{a} = A^{-1}\mathbf{u}$ and that the maximized D is $D_{\text{max}} = \mathbf{u}^\dagger A^{-1}\mathbf{u}$. Show that the value of D_{max} is not affected if $\mathbf{a}$ is defined with an arbitrary normalization factor μ, that is, $\mathbf{a} = \mu A^{-1}\mathbf{u}$.

b. Show that an equivalent problem is the minimization problem:

$$\mathbf{a}^\dagger A\mathbf{a}=\min\,,\quad \text{subject to}\quad \mathbf{u}^\dagger\mathbf{a}=1$$

c. Show that (19.11.1) is a special case of the more general problem of the maximization of the *Rayleigh quotient*:

$$D=\frac{\mathbf{a}^\dagger Q\mathbf{a}}{\mathbf{a}^\dagger A\mathbf{a}}=\max$$

where A, Q are positive-definite Hermitian matrices. Show that the solution of this problem is the eigenvector corresponding to the maximum eigenvalue $\lambda = \lambda_{\text{max}}$ of the generalized eigenvalue problem $Q\mathbf{a} = \lambda A\mathbf{a}$. Explain how this formulation leads to the same solution in the case of (19.11.1).

d. Show that the directivity (19.11.1) of a uniform array ($\mathbf{a} = \mathbf{u}$) is given by the two equivalent forms:

$$D_{\text{unif}}=\frac{|\mathbf{u}^\dagger\mathbf{u}|^2}{\mathbf{u}^\dagger A\mathbf{u}}=\frac{N^2}{N+2\displaystyle\sum_{n=1}^{N-1}(N-|n|)\,\frac{\sin(kdn)}{kdn}}$$

19.7 *Computer Experiment—Optimum Directivity.* Using the matrix formulation of the previous problem, calculate the optimum directivity for an N-element array over the range of spacing values: $0.1 \le d/\lambda \le 2$ and plot it versus d. Carry this out for the values $N = 5, 10, 15$ and place the results on the same graph.

The directivity D of (19.11.1) can be evaluated for any given vector of array weights. Evaluate it for the uniform array $\mathbf{a} = \mathbf{u}$ and plot the results on the same graph as above. You should observe that directivity of the uniform array comes close to that of the optimum one for most (but not all) of the spacings d.

For each d and for the case $N = 15$, calculate the directivities of the array weights $\mathbf{a}$ designed with the MATLAB function `taylor1p` of the next chapter, with sidelobe attenuations of $R = 20$ $R = 30$ dB, and place them on the same graph.

20

Array Design Methods

20.1 Array Design Methods

As we mentioned in Sec. 19.4, the array design problem is essentially equivalent to the problem of designing FIR digital filters in DSP. Following this equivalence, we discuss several array design methods, such as:

1. Schelkunoff's zero placement method
2. Fourier series method with windowing
3. Woodward-Lawson frequency-sampling design
4. Narrow-beam low-sidelobe design methods
5. Multi-beam array design

Next, we establish some common notation. One-dimensional equally-spaced arrays are usually considered *symmetrically* with respect to the origin of the array axis. This requires a slight redefinition of the array factor in the case of *even* number of array elements. Consider an array of N elements at locations x_m along the x-axis with element spacing d. The array factor will be:

$$A(\phi)=\sum_m a_m e^{jk_x x_m}=\sum_m a_m e^{jkx_m\cos\phi}$$

where $k_x = k\cos\phi$ (for polar angle $\theta = \pi/2$.) If N is odd, say $N = 2M+1$, we can define the element locations x_m symmetrically as:

$$x_m = md\,,\qquad m = 0, \pm 1, \pm 2, \dots, \pm M$$

This was the definition we used in Sec. 19.4. The array factor can be written then as a discrete-space Fourier transform or as a spatial z-transform:

$$\boxed{\begin{aligned} A(\psi) &= \sum_{m=-M}^{M} a_m e^{jm\psi} = a_0 + \sum_{m=1}^{M}\Big[a_m e^{jm\psi} + a_{-m}e^{-jm\psi}\Big] \\ A(z) &= \sum_{m=-M}^{M} a_m z^{m} = a_0 + \sum_{m=1}^{M}\big[a_m z^{m} + a_{-m}z^{-m}\big] \end{aligned}} \tag{20.1.1}$$

where $\psi = k_x d = kd\cos\phi$ and $z = e^{j\psi}$. On the other hand, if N is even, say $N = 2M$, in order to have symmetry with respect to the origin, we must place the elements at the *half-integer* locations:

$$x_{\pm m} = \pm\Big(md - \frac{d}{2}\Big) = \pm\Big(m - \frac{1}{2}\Big)d\,, \qquad m = 1, 2, \dots, M$$

The array factor will be now:

$$\boxed{\begin{aligned} A(\psi) &= \sum_{m=1}^{M}\Big[a_m e^{j(m-1/2)\psi} + a_{-m}e^{-j(m-1/2)\psi}\Big] \\ A(z) &= \sum_{m=1}^{M}\Big[a_m z^{m-1/2} + a_{-m}z^{-(m-1/2)}\Big] \end{aligned}} \tag{20.1.2}$$

In particular, if the array weights a_m are symmetric with respect to the origin, $a_m = a_{-m}$, as they are in most design methods, then the array factor can be simplified into the cosine forms:

$$\begin{aligned} A(\psi) &= a_0 + 2\sum_{m=1}^{M} a_m\cos(m\psi)\,, && (N = 2M+1) \\ A(\psi) &= 2\sum_{m=1}^{M} a_m\cos\big((m-1/2)\psi\big)\,, && (N = 2M) \end{aligned} \tag{20.1.3}$$

In *both* the odd and even cases, Eqs. (20.1.1) and (20.1.2) can be expressed as the left-shifted version of a right-sided z-transform:

$$\boxed{A(z) = z^{-(N-1)/2}\tilde{A}(z) = z^{-(N-1)/2}\sum_{n=0}^{N-1}\tilde{a}_n z^n} \tag{20.1.4}$$

where $\mathbf{a} = [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{N-1}]$ is the vector of array weights reindexed to be right-sided. In terms of the original symmetric weights, we have:

$$\begin{aligned} [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{N-1}] &= [a_{-M}, \dots, a_{-1}, a_0, a_1, \dots, a_M]\,, && (N = 2M+1) \\ [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{N-1}] &= [a_{-M}, \dots, a_{-1}, a_1, \dots, a_M]\,, && (N = 2M) \end{aligned} \tag{20.1.5}$$

In time-domain DSP, a factor of z represents a time-advance or left shift. But in the spatial domain, a left shift is represented by z^{-1} because of the opposite sign convention in the definition of the z-transform. Thus, the factor $z^{-(N-1)/2}$ represents a left shift by a distance $(N-1)d/2$, which places the middle of the right-sided array at the origin. For instance, see Examples 19.3.1 and 19.3.2.

The corresponding array factors in ψ-space are related in a similar fashion. Setting $z = e^{j\psi}$, we have:

$$\boxed{A(\psi) = e^{-j\psi(N-1)/2}\tilde{A}(\psi) = e^{-j\psi(N-1)/2}\sum_{n=0}^{N-1}\tilde{a}_n e^{jn\psi}} \tag{20.1.6}$$

Working with $\tilde{A}(\psi)$ is more convenient for programming purposes, as it can be computed by an ordinary DTFT routine, such as that in Ref. [48], $\tilde{A}(\psi) = \text{dtft}(\mathbf{a}, -\psi)$. The phase factor $e^{-j\psi(N-1)/2}$ does not affect the power gain of the array; indeed, we have $|A(\psi)|^2 = |\tilde{A}(\psi)|^2 = |\text{dtft}(\mathbf{a}, -\psi)|^2$.

Some differences arise also for steered array factors. Given a steering phase $\psi_0 = kd\cos\phi_0$, we define the steered array factor as $A'(\psi) = A(\psi - \psi_0)$. Then, we have:

$$A'(\psi) = A(\psi - \psi_0) = e^{-j(\psi-\psi_0)(N-1)/2}\tilde{A}(\psi - \psi_0) = e^{-j\psi(N-1)/2}\tilde{A}'(\psi)$$

It follows that the steered version of $\tilde{A}(\psi)$ will be:

$$\boxed{\tilde{A}'(\psi) = e^{j\psi_0(N-1)/2}\tilde{A}(\psi - \psi_0)} \tag{20.1.7}$$

which implies for the weights:

$$\boxed{\tilde{a}'_n = \tilde{a}_n e^{-j\psi_0(n-(N-1)/2)}}\,, \qquad n = 0, 1, \dots, N-1 \tag{20.1.8}$$

This simply means that the progressive phase is measured with respect to the middle of the array. Again, the common phase factor $e^{j\psi_0(N-1)/2}$ is usually unimportant. One case where it is important is the case of *multiple* beams steered towards different angles; these are discussed in Sec. 20.14. In the *symmetric* notation, the steered weights are as follows:

$$\begin{aligned} a'_m &= a_m e^{-jm\psi_0}\,, \quad m = 0, \pm 1, \pm 2, \dots, \pm M\,, \quad (N = 2M+1) \\ a'_{\pm m} &= a_{\pm m}e^{\mp j(m-1/2)\psi_0}\,, \quad m = 1, 2, \dots, M\,, \quad (N = 2M) \end{aligned} \tag{20.1.9}$$

The MATLAB functions `scan` and `steer` perform the desired progressive phasing of the weights according to Eq. (20.1.8). Their usage is as follows:

```
ascan = scan(a, psi0);          % scan array with given scanning phase ψ0
asteer = steer(d, a, ph0);      % steer array towards given angle φ0
```

Example 20.1.1: For the cases $N = 7$ and $N = 6$, we have $M = 3$. The symmetric and right-sided array weights will be related as follows:

$$\mathbf{a} = [\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4, \tilde{a}_5, \tilde{a}_6] = [a_{-3}, a_{-2}, a_{-1}, a_0, a_1, a_2, a_3]$$

$$\mathbf{a} = [\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4, \tilde{a}_5] = [a_{-3}, a_{-2}, a_{-1}, a_1, a_2, a_3]$$

For $N = 7$ we have $(N-1)/2 = 3$, and for $N = 6$, $(N-1)/2 = 5/2$. Thus, the array locations along the x-axis will be:

$$x_m = \{-3d,\ -2d,\ -d,\ 0,\ d,\ 2d,\ 3d\}$$

$$x_m = \{-\frac{5}{2}d,\ -\frac{3}{2}d,\ -\frac{1}{2}d,\ \frac{1}{2}d,\ \frac{3}{2}d,\ \frac{5}{2}d\}$$

Eq. (20.1.4) reads as follows in the two cases:

$$\begin{aligned} A(z) &= a_{-3}z^{-3} + a_{-2}z^{-2} + a_{-1}z^{-1} + a_0 + a_1 z + a_2 z^2 + a_3 z^3 \\ &= z^{-3}\big[a_{-3} + a_{-2}z + a_{-1}z^2 + a_0 z^3 + a_1 z^4 + a_2 z^5 + a_3 z^6\big] = z^{-3}\tilde{A}(z) \\ A(z) &= a_{-3}z^{-5/2} + a_{-2}z^{-3/2} + a_{-1}z^{-1/2} + a_1 z^{1/2} + a_2 z^{3/2} + a_3 z^{5/2} \\ &= z^{-5/2}\big[a_{-3} + a_{-2}z + a_{-1}z^2 + a_1 z^3 + a_2 z^4 + a_3 z^5\big] = z^{-5/2}\tilde{A}(z) \end{aligned}$$

If the arrays are steered, the weights pick up the progressive phases:

$$
\begin{aligned}
&[a_{-3}e^{j3\psi_0},\ a_{-2}e^{j2\psi_0},\ a_{-1}e^{j\psi_0},\ a_0,\ a_1e^{-j\psi_0},\ a_2e^{-j2\psi_0},\ a_3e^{-j3\psi_0}]\\
&\quad = e^{j3\psi_0}[a_{-3},\ a_{-2}e^{-j\psi_0},\ a_{-1}e^{-2j\psi_0},\ a_0e^{-3j\psi_0},\ a_1e^{-4j\psi_0},\ a_2e^{-j5\psi_0},\ a_3e^{-j6\psi_0}]\\
&[a_{-3}e^{j5\psi_0/2},\ a_{-2}e^{j3\psi_0/2},\ a_{-1}e^{j\psi_0/2},\ a_1e^{-j\psi_0/2},\ a_2e^{-j3\psi_0/2},\ a_3e^{-j5\psi_0/2}]\\
&\quad = e^{j5\psi_0/2}[a_{-3},\ a_{-2}e^{-j\psi_0},\ a_{-1}e^{-2j\psi_0},\ a_1e^{-3j\psi_0},\ a_2e^{-j4\psi_0},\ a_3e^{-j5\psi_0}]
\end{aligned}
$$

where $\psi_0 = kd\cos\phi_0$ is the steering phase. □

Example 20.1.2: The uniform array of Sec. 19.7, was defined as a right-sided array. In the present notation, the weights and array factor are:

$$\mathbf{a} = [\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_{N-1}] = \frac{1}{N}[1, 1, \dots, 1], \qquad \tilde{A}(z) = \frac{1}{N}\frac{z^N - 1}{z - 1}$$

Using Eq. (20.1.4), the corresponding symmetric array factor will be:

$$A(z) = z^{-(N-1)/2}\tilde{A}(z) = z^{-(N-1)/2}\frac{1}{N}\frac{z^N - 1}{z - 1} = \frac{1}{N}\frac{z^{N/2} - z^{-N/2}}{z^{1/2} - z^{-1/2}}$$

Setting $z = e^{j\psi}$, we obtain

$$A(\psi) = \frac{\sin\left(\dfrac{N\psi}{2}\right)}{N\sin\left(\dfrac{\psi}{2}\right)} \tag{20.1.10}$$

which also follows from Eqs. (19.7.3) and (20.1.6). □

20.2 Schelkunoff's Zero Placement Method

The array factor of an N-element array is a polynomial of degree $N - 1$ and therefore it has $N - 1$ zeros:

$$\tilde{A}(z) = \sum_{n=0}^{N-1}\tilde{a}_n z^n = (z - z_1)(z - z_2)\cdots(z - z_{N-1})\tilde{a}_{N-1} \tag{20.2.1}$$

By proper placement of the zeros on the z-plane, a desired array factor can be designed. Schelkunoff's paper of more than 45 years ago [1067] discusses this and the Fourier series methods.

As an example consider the uniform array that has zeros equally spaced around the unit circle at the N-th roots of unity, that is, at $z_i = e^{j\psi_i}$, where $\psi_i = 2\pi i/N$, $i = 1, 2, \dots, N - 1$. The index $i = 0$ is excluded as $z = 1$ or $\psi = 0$ corresponds to the mainlobe peak of the array. Depending on the element spacing d, it is possible that not all of these zeros lie within the visible region and, therefore, they may not correspond to actual nulls in the angular pattern. This happens when $d < \lambda/2$ for a broadside array, which has a visible region that covers less than the full unit circle, $\psi_{\text{vis}} = 2kd < 2\pi$.

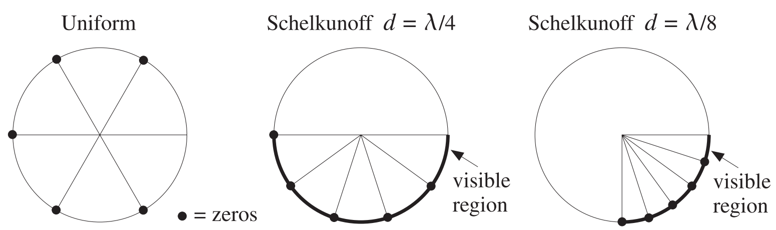


Fig. 20.2.1 Endfire array zeros and visible regions for $N = 6$, and $d = \lambda/4$ and $d = \lambda/8$.

Schelkunoff's design idea was to place all $N - 1$ zeros of the array within the visible region, for example, by equally spacing them within it. Fig. 20.2.1 shows the visible regions and array zeros for a six-element *endfire* array with element spacings $d = \lambda/4$ and $d = \lambda/8$.

The visible region is determined by Eq. (19.9.5). For an endfire ($\phi_0 = 0$) array with $d = \lambda/4$ or $kd = \pi/2$, the steered wavenumber will be $\psi' = kd(\cos\phi - \cos\phi_0) = (\cos\phi - 1)\pi/2$ and the corresponding visible region, $-\pi \le \psi' \le 0$. Similarly, when $d = \lambda/8$ or $kd = \pi/4$, we have $\psi' = (\cos\phi - 1)\pi/4$ and visible region, $-\pi/2 \le \psi' \le 0$.

The uniform array has five zeros. When $d = \lambda/4$, only three of them lie in the visible region, and when $d = \lambda/8$ only one of them does. By contrast Schelkunoff's design method places all five zeros within the visible regions.

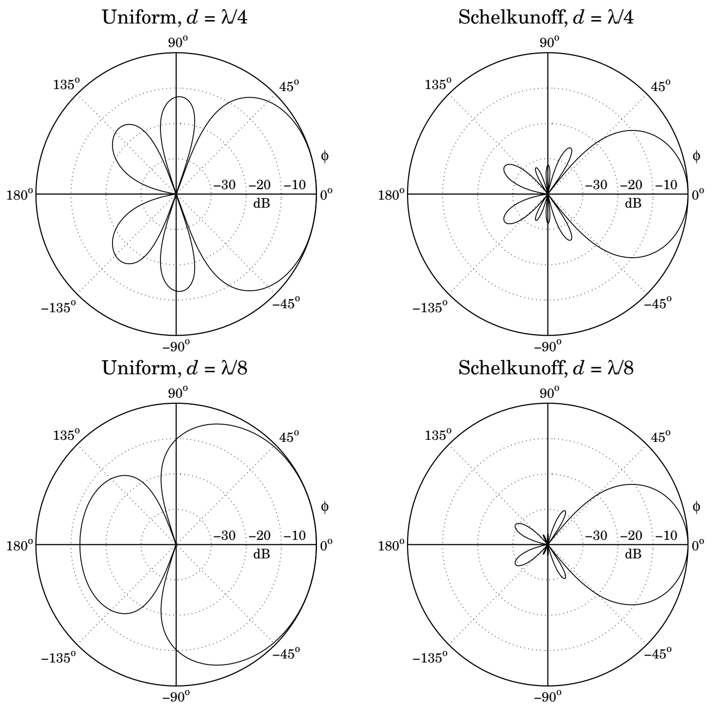
Fig. 20.2.2 shows the gains of the two cases and compares them to the gains of the corresponding uniform array. The presence of more zeros in the visible regions results in a narrower mainlobe and smaller sidelobes.

The angular nulls corresponding to the zeros that lie in the visible region may be observed in these graphs for both the uniform and Schelkunoff designs.

Because the visible region is in both cases $-2kd \le \psi' \le 0$, the five zeros are chosen as $z_i = e^{j\psi_i}$, where $\psi_i = -2kdi/5$, $i = 1, 2, \dots, 5$. The array weights can be obtained by expanding the zero factors of Eq. (20.2.1). The following MATLAB statements will perform and plot the design:

```
d=1/4; kd=2*pi*d;
i = 1:5;
psi = -2*kd*i/5;
zi = exp(j*psi);
a = fliplr(poly(zi));
a = steer(d, a, 0);
[g, ph] = array(d, a, 400);
dbz(ph, g, 45, 40);
```

The function `poly` computes the expansion coefficients. But because it lists them from the higher coefficient to the lowest one, that is, from z^{N-1} to z^0, it is necessary to reverse the vector by `fliplr`. When the weight vector is symmetric with respect to its middle, such reversal is not necessary.



Uniform, $d = \lambda/4$

Schelkunoff, $d = \lambda/4$

Uniform, $d = \lambda/8$

Schelkunoff, $d = \lambda/8$

Fig. 20.2.2 Gain of six-element endfire array with $d = \lambda/4$ and $d = \lambda/8$.

20.3 Fourier Series Method with Windowing

The Fourier series design method is identical to the same method in DSP for designing FIR digital filters [47,48]. The method is based on the inverse discrete-space Fourier transforms of the array factor.

Eqs. (20.1.1) and (20.1.2) may be thought of as the truncated or *windowed* versions of the corresponding infinite Fourier series. Assuming an infinite and convergent series, we have for the "odd" case:

$$\boxed{A(\psi)= a_0 + \sum_{m=1}^{\infty}\Big[a_m e^{jm\psi} + a_{-m}e^{-jm\psi}\Big]} \tag{20.3.1}$$

Then, the corresponding inverse transform will be:

$$\boxed{a_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\psi)\,e^{-jm\psi}\,d\psi}\,, \qquad m = 0, \pm 1, \pm 2, \dots \tag{20.3.2}$$

Similarly, in the "even" case we have:

$$\boxed{A(\psi)= \sum_{m=1}^{\infty}\Big[a_m e^{j(m-1/2)\psi} + a_{-m}e^{-j(m-1/2)\psi}\Big]} \tag{20.3.3}$$

with inverse transform:

$$\boxed{a_{\pm m} = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\psi)\,e^{\mp j(m-1/2)\psi}\,d\psi}\,, \qquad m = 1, 2, \dots \tag{20.3.4}$$

In general, a desired array factor requires an infinite number of coefficients a_m to be represented exactly. Keeping only a finite number of coefficients in the Fourier series introduces unwanted ripples in the desired response, known as the Gibbs phenomenon [47,48]. Such ripples can be minimized using an appropriate window, but at the expense of wider transition regions.

The Fourier series method may be summarized as follows. Given a desired response, say $A_d(\psi)$, pick an odd or even window length, for example $N = 2M + 1$, and calculate the N ideal weights by evaluating the inverse transform:

$$a_d(m)= \frac{1}{2\pi}\int_{-\pi}^{\pi} A_d(\psi)\,e^{-jm\psi}\,d\psi\,, \qquad m = 0, \pm 1, \dots, \pm M \tag{20.3.5}$$

then, the final weights are obtained by windowing with a length-N window $w(m)$:

$$a(m)= w(m)\,a_d(m)\,, \qquad m = 0, \pm 1, \dots, \pm M \tag{20.3.6}$$

This method is convenient only when the required integral (20.3.5) can be done exactly, as when $A_d(\psi)$ has a simple shape such as an ideal lowpass filter. For arbitrarily shaped $A_d(\psi)$ one must evaluate the integrals approximately using an inverse DFT as is done in the Woodward- Lawson frequency-sampling design method discussed in Sec. 20.5.

In addition, the method requires that $A_d(\psi)$ be specified over one complete Nyquist interval, $-\pi \le \psi \le \pi$, regardless of whether the visible region $\psi_{\text{vis}} = 2kd$ is more or less than one Nyquist period.

20.4 Sector Beam Array Design

As an example of the Fourier series method, we discuss the design of an array with angular pattern confined into a desired angular sector.

First, we consider the design in ψ-space of an ideal *bandpass* array factor centered at wavenumber ψ_0 with bandwidth of $2\psi_b$. We will see later how to map these specifications into an actual angular sector. The ideal bandpass response is defined over $-\pi \le \psi \le \pi$ as follows:

$$A_{\text{BP}}(\psi)= \begin{cases} 1, & \psi_0 - \psi_b \le \psi \le \psi_0 + \psi_b \\ 0, & \text{otherwise} \end{cases}$$

For the odd case, the corresponding ideal weights are obtained from Eq. (20.3.2):

$$a_{\rm BP}(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}A_{\rm BP}(\psi)e^{-jm\psi}\,d\psi=\frac{1}{2\pi}\int_{\psi_0-\psi_b}^{\psi_0+\psi_b}1\cdot e^{-jm\psi}\,d\psi$$

which gives:

$$a_{\rm BP}(m)=e^{-jm\psi_0}\frac{\sin(\psi_b m)}{\pi m}\,,\qquad m=0,\pm1,\pm2,\dots \tag{20.4.1}$$

This problem is equivalent to designing an ideal *lowpass* response with cutoff frequency ψ_b and then translating it by $A_{\rm BP}(\psi)=A_{\rm LP}(\psi')=A_{\rm LP}(\psi-\psi_0)$, where $\psi'=\psi-\psi_0$. The lowpass response is defined as:

$$A_{\rm LP}(\psi')=\begin{cases}1, & -\psi_b\le\psi'\le\psi_b\\ 0, & \text{otherwise}\end{cases}$$

and its ideal weights are:

$$a_{\rm LP}(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}A_{\rm LP}(\psi')e^{-jm\psi'}\,d\psi'=\frac{1}{2\pi}\int_{-\psi_b}^{\psi_b}1\cdot e^{-jm\psi'}\,d\psi'=\frac{\sin(\psi_b m)}{\pi m}$$

Thus, as expected, the ideal weights for the bandpass and lowpass designs are related by a scanning phase: $a_{\rm BP}(m)=e^{-jm\psi_0}a_{\rm LP}(m)$.

A more realistic design of the bandpass response is to prescribe "brickwall" specifications, that is, defining a passband range over which the response is essentially flat and a stopband range over which the response is essentially zero. These ranges are defined by the bandedge frequencies ψ_p and ψ_s, such that the passband is $|\psi-\psi_0|\le\psi_p$ and the stopband $|\psi-\psi_0|\ge\psi_s$. The specifications of the equivalent lowpass response are shown in Fig. 20.4.1.

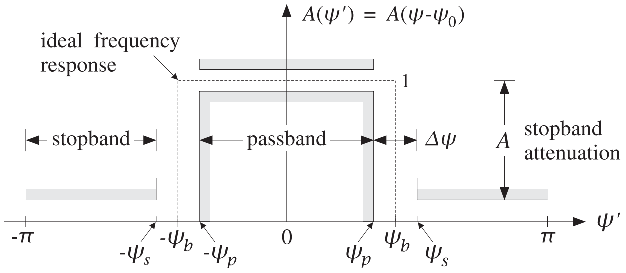


Fig. 20.4.1 Specifications of equivalent lowpass response.

Over the stopband, the attenuation is required to be greater than a *minimum* value, say A dB. The attenuation over the passband need not be specified, because the window method always results in extremely flat passbands for reasonable values of A, e.g., for $A>35$ dB. Indeed, the maximum passband attenuation is related to A by the approximate formula $A_{\rm pass}=17.4\delta$ dB, where $\delta=10^{-A/20}$ (see Ref. [48].)

Most windows do not allow a user-defined choice for the stopband attenuation. For example, the Hamming window has $A=54$ dB and the rectangular window $A=21$ dB.

The *Kaiser window* is the best and simplest of a small class of windows that allow a *variable* choice for A.

Thus, the design specifications are the quantities $\{\psi_p,\psi_s,A\}$. Alternatively, we can take them to be $\{\psi_p,\Delta\psi,A\}$, where $\Delta\psi=\psi_s-\psi_p$ is the transition width. We prefer the latter choice. The design steps for the bandpass response using the Kaiser window are summarized below:

1. From the stopband attenuation A, calculate the so-called D-factor of the window (similar to the broadening factor):

$$D=\begin{cases}\dfrac{A-7.95}{14.36}\,, & \text{if } A>21\\ 0.922, & \text{if } A\le 21\end{cases} \tag{20.4.2}$$

   and the window's shape parameter α:

$$\alpha=\begin{cases}0.1102(A-8.7)\,, & \text{if } A\ge 50\\ 0.5842(A-21)^{0.4}+0.07886(A-21)\,, & \text{if } 21<A<50\\ 0, & \text{if } A\le 21\end{cases} \tag{20.4.3}$$

2. From the transition width $\Delta\psi$, calculate the length of the window by choosing the smallest *odd* integer $N=2M+1$ that satisfies:

$$\boxed{\Delta\psi=\frac{2\pi D}{N-1}} \tag{20.4.4}$$

   Alternatively, if N is given, calculate the transition width $\Delta\psi$.

3. Calculate the samples of the Kaiser window:

$$w(m)=\frac{I_0\big(\alpha\sqrt{1-m^2/M^2}\big)}{I_0(\alpha)}\,,\qquad m=0,\pm1,\dots,\pm M \tag{20.4.5}$$

   where $I_0(x)$ is the modified Bessel function of first kind and zeroth order.

4. Calculate the ideal cutoff frequency ψ_b by taking it to be at the middle between the passband and stopband frequencies:

$$\psi_b=\frac{1}{2}(\psi_p+\psi_s)=\psi_p+\frac{1}{2}\Delta\psi \tag{20.4.6}$$

5. Calculate the final windowed array weights from $a(m)=w(m)a_{\rm BP}(m)$:

$$\boxed{a(m)=w(m)e^{-jm\psi_0}\frac{\sin(\psi_b m)}{\pi m}}\,,\qquad m=0,\pm1,\dots,\pm M \tag{20.4.7}$$

Next, we use the above bandpass design in ψ-space to design an array with an angular sector response in ϕ-space. The ideal array will have a pattern that is uniformly flat over an angular sector $[\phi_1,\phi_2]$:

$$A(\phi)=\begin{cases}1, & \phi_1\le\phi\le\phi_2\\ 0, & \text{otherwise}\end{cases}$$

Alternatively, we can define the sector by means of its center angle and its width, $\phi_c=(\phi_1+\phi_2)/2$ and $\phi_b=\phi_2-\phi_1$. Thus, we have the equivalent definitions of the angular sector:

$$\begin{aligned}\phi_c&=\frac{1}{2}(\phi_1+\phi_2)\\ \phi_b&=\phi_2-\phi_1\end{aligned}\quad\Leftrightarrow\quad\begin{aligned}\phi_1&=\phi_c-\frac{1}{2}\phi_b\\ \phi_2&=\phi_c+\frac{1}{2}\phi_b\end{aligned}\tag{20.4.8}$$

For a practical design, we may take $[\phi_1,\phi_2]$ to represent the *passband* of the response and assume an angular stopband with attenuation of at least A dB that begins after a small angular transition width $\Delta\phi$ on either side of the passband.

In filter design, the stopband attenuation and the transition width are used to determine the window length N. But in the array problem, because we are usually limited in the number N of available array elements, we must assume that N is given and determine the transition width $\Delta\phi$ from A and N.

Thus, our design specifications are the quantities $\{\phi_1,\phi_2,N,A\}$, or alternatively, $\{\phi_c,\phi_b,N,A\}$. These specifications must be mapped into equivalent ones in ψ-space using the steered wavenumber $\psi'=kd(\cos\phi-\cos\phi_0)$.

We require that the angular passband $[\phi_1,\phi_2]$ be mapped onto the lowpass passband $[-\psi_p,\psi_p]$ in ψ'-space. Thus, we have the conditions:

$$\begin{aligned}\psi_p&=kd\cos\phi_1-\psi_0\\ -\psi_p&=kd\cos\phi_2-\psi_0\end{aligned}$$

They may be solved for ψ_p and ψ_0 as follows:

$$\begin{aligned}\psi_p&=\frac{1}{2}kd(\cos\phi_1-\cos\phi_2)\\ \psi_0&=\frac{1}{2}kd(\cos\phi_1+\cos\phi_2)\end{aligned}\tag{20.4.9}$$

Using Eq. (20.4.8) and some trigonometry, we have equivalently:

$$\boxed{\begin{aligned}\psi_p&=kd\sin(\phi_c)\sin\Big(\frac{\phi_b}{2}\Big)\\ \psi_0&=kd\cos(\phi_c)\cos\Big(\frac{\phi_b}{2}\Big)\end{aligned}}\tag{20.4.10}$$

Setting $\psi_0=kd\cos\phi_0$, we find the effective steering angle ϕ_0:

$$\cos\phi_0=\cos(\phi_c)\cos\Big(\frac{\phi_b}{2}\Big)\quad\Rightarrow\quad\phi_0=\operatorname{acos}\big(\cos(\phi_c)\cos(\phi_b/2)\big)\tag{20.4.11}$$

Note that ϕ_0 is not equal to ϕ_c, except for very narrow widths ϕ_b.

The design procedure is then completed as follows. Given the attenuation A, we calculate the window parameters D,α from Eqs. (20.4.2) and (20.4.3). Since N is given, we calculate the transition width $\Delta\psi$ directly from Eq. (20.4.4). Then, the ideal lowpass frequency ψ_b is calculated from Eq. (20.4.6), that is,

$$\psi_b=\psi_p+\frac{1}{2}\Delta\psi=kd\sin(\phi_c)\sin\Big(\frac{\phi_b}{2}\Big)+\frac{\pi D}{N-1}\tag{20.4.12}$$

Finally, the array weights are obtained from Eq. (20.4.7). The transition width $\Delta\phi$ can be approximated by linearizing $\psi=kd\cos\phi$ around ϕ_1, or around ϕ_2, or around ϕ_c. We prefer the latter choice, giving:

$$\Delta\phi=\frac{\Delta\psi}{kd\sin\phi_c}=\frac{2\pi D}{kd(N-1)\sin\phi_c}\tag{20.4.13}$$

The design method can be extended to the case of even $N=2M$. The integral (20.3.4) can still be done exactly. The Kaiser window expression (20.4.5) remains the same for $m=\pm1,\pm2,\dots,\pm M$. We note the symmetry $w(-m)=w(m)$. After windowing and scanning with ψ_0, we get the final designed weights:

$$\boxed{a(\pm m)=w(m)e^{\mp j(m-1/2)\psi_0}\,\frac{\sin\big(\psi_b(m-1/2)\big)}{\pi(m-1/2)}},\quad m=1,2,\dots,M\tag{20.4.14}$$

The MATLAB function `sector` implements the above design steps for either even or odd N. Its usage is as follows:

```
[a, dph] = sector(d, ph1, ph2, N, A);        % A=stopband attenuation in dB
```

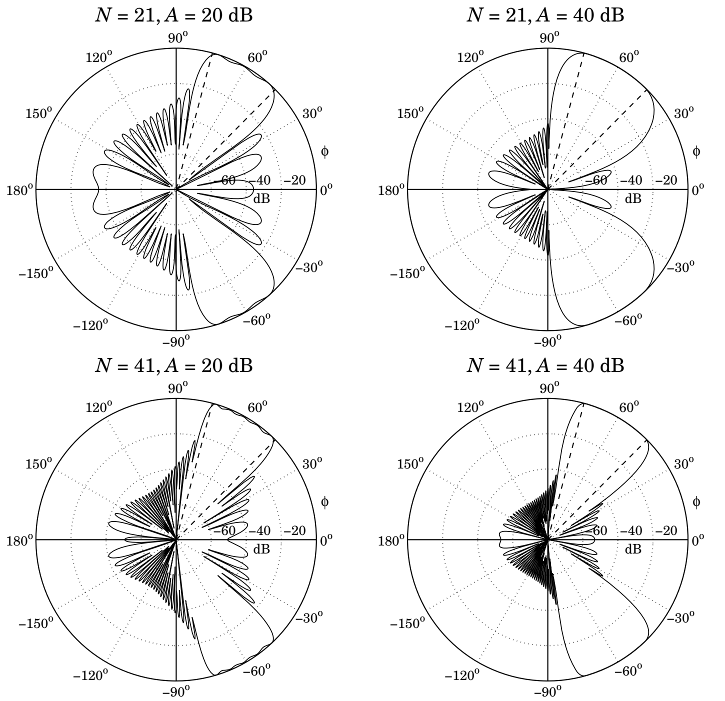
Fig. 20.4.2 shows four design examples having sector $[\phi_1,\phi_2]=[45^o,75^o]$, or center $\phi_c=60^o$ and width $\phi_b=30^o$. The number of array elements was $N=21$ and $N=41$, with half-wavelength spacing $d=\lambda/2$. The stopband attenuations were $A=20$ and $A=40$ dB. The two cases with $A=20$ dB are equivalent to using the rectangular window. They have visible Gibbs ripples in their passband. Some typical MATLAB code for generating these graphs is as follows:

```
d=0.5; ph1=45; ph2=75; N=21; A=20;
[a, dph] = sector(d, ph1, ph2, N, A);
[g, ph] = array(d, a, 400);
dbz(ph,g, 30, 80);
addray(ph1, '--'); addray(ph2, '--');
```

The basic design tradeoff is between N and A and is captured by Eq. (20.4.4). Because D is linearly increasing with A, the transition width will increase with A and decrease with N. As A increases, the passband exhibits no Gibbs ripples but at the expense of larger transition width.

20.5 Woodward-Lawson Frequency-Sampling Design

As we mentioned earlier, the Fourier series method is feasible only when the inverse transform integrals (20.3.2) and (20.3.4) can be done exactly. If not, we may use the


$N=21, A=20$ dB

$N=21, A=40$ dB

$N=41, A=20$ dB

$N=41, A=40$ dB

Fig. 20.4.2 Angular sector array design with the Kaiser window.

frequency-sampling design method of DSP [47,48]. In the array context, the method is referred to as the *Woodward-Lawson* method.

For an N-element array, the method is based on performing an inverse N-point DFT. It assumes that N samples of the desired array factor $A(\psi)$ are available, that is, $A(\psi_i)$, $i=0,1,\dots,N-1$, where ψ_i are the N DFT frequencies:

$$\boxed{\psi_i=\frac{2\pi i}{N}}\,,\qquad i=0,1,\dots,N-1,\qquad \text{(DFT frequencies)} \tag{20.5.1}$$

The frequency samples $A(\psi_i)$ are related to the array weights via the *forward* N-point DFT's obtained by evaluating Eqs. (20.1.1) and (20.1.2) at the N DFT frequencies:

$$\begin{aligned} A(\psi_i)&=a_0+\sum_{m=1}^{M}\Big[a_m e^{jm\psi_i}+a_{-m}e^{-jm\psi_i}\Big]\,, && (N=2M+1)\\ A(\psi_i)&=\sum_{m=1}^{M}\Big[a_m e^{j(m-1/2)\psi_i}+a_{-m}e^{-j(m-1/2)\psi_i}\Big]\,, && (N=2M) \end{aligned} \tag{20.5.2}$$

where ψ_i are given by Eq. (20.5.1). The corresponding *inverse* N-point DFT's are as follows. For odd $N=2M+1$,

$$\boxed{a_m=\frac{1}{N}\sum_{i=0}^{N-1}A(\psi_i)e^{-jm\psi_i}}\,,\qquad m=0,\pm1,\pm2,\dots,\pm M \tag{20.5.3}$$

and for even $N=2M$,

$$\boxed{a_{\pm m}=\frac{1}{N}\sum_{i=0}^{N-1}A(\psi_i)e^{\mp j(m-1/2)\psi_i}}\,,\qquad m=1,2,\dots,M \tag{20.5.4}$$

There is an alternative definition of the N DFT frequencies ψ_i for which the forms of the forward and inverse DFT's, Eqs. (20.5.2)–(20.5.4), remain the same. For either even or odd N, we define:

$$\boxed{\psi_i=\frac{2\pi(i-K)}{N}}\,,\qquad \text{(alternative DFT frequencies)} \tag{20.5.5}$$

where $i=0,1,\dots,N-1$ and $K=(N-1)/2$.

This definition makes a difference only for even N, in which case the index $i-K$ takes on all the *half-integer* values in the *symmetric* interval $[-K,K]$. For odd N, Eq. (20.5.5) amounts to a re-indexing of Eq. (20.5.1), with $i-K$ taking values now over the symmetric integer interval $[-K,K]$.

For both the standard and the alternative sets, the N complex numbers $z_i=e^{j\psi_i}$ are equally spaced around the unit circle. For odd N, they are the N-th roots of unity, that is, the solutions of the equation $z^N=1$. For the alternative set with even N, they are the N solutions of the equation $z^N=-1$.

The alternative set is usually preferred in array processing. In DSP, it leads to the discrete cosine transform. The MATLAB function `woodward` implements the inverse DFT operations (20.5.3) and (20.5.4), for either the standard or the alternative definition of ψ_i. Its usage is as follows:

```
a = woodward(A, alt);          % alt=0,1 for standard or alternative
```

The frequency-sampling array design method is summarized as follows: Given a set of N frequency response values $A(\psi_i)$, $i=0,1,\dots,N-1$, calculate the N array weights $a(m)$ using the inverse DFT formulas (20.5.3) or (20.5.4). Then, replace the weights by their windowed versions using any symmetric length-N window. The final expressions for the windowed weights are, for odd $N=2M+1$,

$$\boxed{a(m)=w(m)\frac{1}{N}\sum_{i=0}^{N-1}A(\psi_i)e^{-jm\psi_i}}\,,\qquad m=0,\pm1,\pm2,\dots,\pm M \tag{20.5.6}$$

and for even $N=2M$,

$$\boxed{a(\pm m)=w(\pm m)\frac{1}{N}\sum_{i=0}^{N-1}A(\psi_i)e^{\mp j(m-1/2)\psi_i}}\,,\qquad m=1,2,\dots,M \tag{20.5.7}$$

As an example, consider the design of a sector beam with edges at $\phi_1 = 45^{\rm o}$ and $\phi_2 = 75^{\rm o}$. Thus, the beam is centered at $\phi_c = 60^{\rm o}$ and has width $\phi_b = 30^{\rm o}$.

As ϕ ranges over $[\phi_1, \phi_2]$, the wavenumber $\psi = kd\cos\phi$ will range over $kd\cos\phi_2 \le \psi \le kd\cos\phi_1$. For all DFT frequencies ψ_i that lie in this interval, we set $A(\psi_i)= 1$, otherwise, we set $A(\psi_i)= 0$. Assuming the alternative definition for ψ_i, we have the passband condition:

$$kd\cos\phi_2 \le \frac{2\pi(i-K)}{N} \le kd\cos\phi_1$$

Setting $kd = 2\pi d/\lambda$ and solving for the DFT index $i-K$, we find:

$$j_1 \le i-K \le j_2\,, \qquad \text{where} \quad j_1 = \frac{Nd}{\lambda}\cos\phi_2\,, \quad j_2 = \frac{Nd}{\lambda}\cos\phi_1$$

This range determines the DFT indices i for which $A(\psi_i)= 1$. The inverse DFT summation over i will then be restricted over this subset of i's. Fig. 20.5.1 shows the response of a 20-element array with half-wavelength spacing, $d = \lambda/2$, designed with a rectangular and a Hamming window. The MATLAB code for generating the right graph was as follows:

```
d=0.5; N=20; ph1=45; ph2=75; alt=1; K=(N-1)/2;
j1 = N*d*cos(ph2*pi/180);
j2 = N*d*cos(ph1*pi/180);
i = (0:N-1);                                % DFT index
j = i - alt*K;                              % alternative DFT index
A = (j>=j1)&(j<=j2);                        % equals 1, if j1 ≤ j ≤ j2, and 0, otherwise
a = woodward(A, alt);                       % inverse DFT
w = 0.54 - 0.46*cos(2*pi*i/(N-1));          % Hamming window
awind = a .* w;                             % windowed weights
[g,ph] = array(0.5, awind, 400);            % array gain
dbz(ph, g, 30, 80);
addray(ph1,'--'); addray(ph2,'--');
```

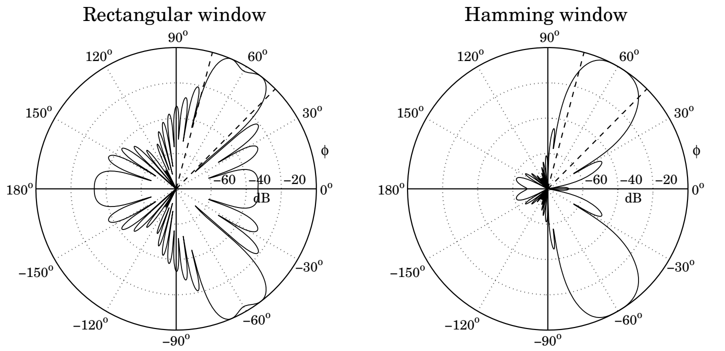


Fig. 20.5.1 Angular sector array design with Woodward-Lawson method.

The sidelobes of the Hamming window are down approximately at the expected 54-dB level (they reach 54 dB for larger N.) The design is comparable to that of Fig. 20.4.2.

The power of this method lies in the ability to specify *any* shape for the array factor through its frequency samples. The method works well for half-wavelength spacing $d = \lambda/2$, because all N DFT frequencies ψ_i lie within the visible region, which coincides in this case with the full Nyquist interval, $-\pi \le \psi \le \pi$.

As another example, we consider the design of an array with a secant-squared gain pattern, which is relevant in air search radars as discussed in Sec. 15.11. We consider an array of N elements along the z-direction with half-wavelength spacing $d = \lambda/2$. The corresponding wavenumber ψ will be $\psi = k_z d$, or

$$\psi = kd\cos\theta$$

The design of the secant-squared gain pattern requires that the array factor itself have a secant dependence. Indeed,

$$g(\theta)= |A(\psi)|^2 = \frac{K}{\cos^2\theta} \quad\Rightarrow\quad |A(\psi)| = \frac{K^{1/2}}{|\cos\theta|}$$

Because the secant pattern is defined only up to an angle $\theta_{\rm max}$, we may define the theoretical array factor in the normalized form:

$$A(\theta)= \begin{cases} \dfrac{\cos\theta_{\rm max}}{\cos\theta}\,, & \text{if } 0 \le \theta \le \theta_{\rm max} \\ 1, & \text{if } \theta_{\rm max} < \theta \le 90^{\rm o} \end{cases} \tag{20.5.8}$$

As θ varies over $[0, \theta_{\rm max}]$, the wavenumber $\psi = kd\cos\theta$ will vary over $[\psi_{\rm max}, kd]$, where $\psi_{\rm max} = kd\cos\theta_{\rm max}$. Because $d = \lambda/2$, we have $kd = \pi$ and the ψ-range becomes $[\psi_{\rm max}, \pi]$. Noting that $\cos\theta_{\rm max}/\cos\theta = \psi_{\rm max}/\psi$, we can rewrite Eq. (20.5.8) in terms of ψ:

$$A(\psi)= \begin{cases} \dfrac{\psi_{\rm max}}{\psi}\,, & \text{if } \psi_{\rm max} \le \psi \le \pi \\ 1, & \text{if } 0 \le \psi < \psi_{\rm max} \end{cases} \tag{20.5.9}$$

We symmetrize $A(-\psi)= A(\psi)$ to cover the entire 2π Nyquist interval in ψ. Evaluating Eq. (20.5.9) at the N DFT frequencies $\psi_i = 2\pi i/N$, we obtain the array weights by doing an inverse DFT and then windowing the array coefficients with a Hamming window. Fig. 20.5.2 shows a design case with $N = 21$ and $\theta_{\rm max} = 70^{\rm o}$. The figure compares the Hamming and rectangular window designs to the exact expression (20.5.8). The details of the design are indicated in the MATLAB code:

```
N=21; K=(N-1)/2; d=0.5; thmax=70;

psmax = 2*pi*d * cos(thmax*pi/180);

Ai = ones(1,K+1);
psi = 2*pi*(0:K)/N;                                          % half of DFT frequencies

j = find(psi);                                               % non-zero ψ's
Ai(j) = psmax*(psi(j)>=psmax)./psi(j) + (psi(j)<psmax);      % half of the DFT values

Ai = [Ai, Ai(K:-1:1)];                                       % all the DFT values

a  = woodward(Ai, 0) / N;                                    % inverse DFT with alt=0
```

```
aw = a .* (0.54 - 0.46*cos(2*pi*(0:N-1)/(N-1)));          % Hamming

th = (0:200) * 90 / 200;
ps = 2*pi*d * cos(th*pi/180);

A  = abs(dtft(a, -ps));                                   % rectangular design
Aw = abs(dtft(aw,-ps));                                   % Hamming design
A0 = psmax*(ps>=psmax)./ps + (ps<psmax);                  % exact pattern
```

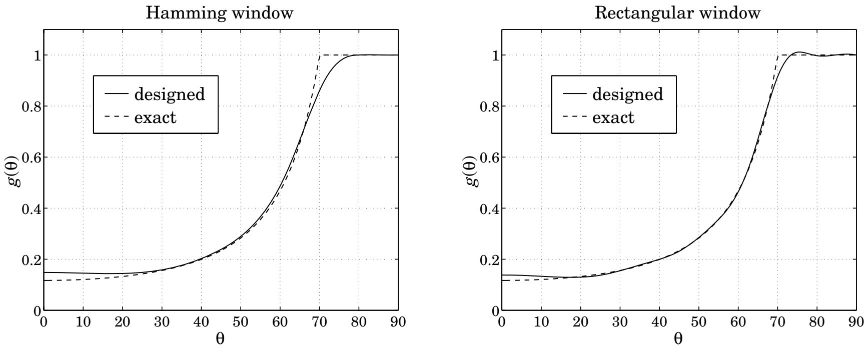


Fig. 20.5.2 Woodward-Lawson design of secant-squared array gain.

20.6 Discretization of Continuous Line Sources

One-dimensional arrays may be thought of as arising from the *spatial sampling* of continuous line current distributions. Consider, for example, a current $I(x)$ flowing along the x-axis. Its current density is $J_x(x,y,x)=I(x)\delta(y)\delta(z)$, where the delta functions confine the current on the x-axis. The corresponding radiation vector will have only an x-component:

$$\begin{aligned} F_x(k_x,k_y,k_z) &= \int J_x(x,y,z)\,e^{jk_x x+jk_y y+jk_z z}\,dx\,dy\,dz \\ &= \int I(x)\,\delta(y)\,\delta(z)\,e^{jk_x x+jk_y y+jk_z z}\,dx\,dy\,dz = \int_{-\infty}^{\infty} I(x)\,e^{jk_x x}dx \end{aligned}$$

Thus, $F_x(k_x)$ depends only on the k_x wavevector component and is the spatial Fourier transform of the line current $I(x)$:

$$F_x(k_x)=\int_{-\infty}^{\infty} I(x)\,e^{jk_x x}dx \tag{20.6.1}$$

In spherical coordinates, k_x is given by $k_x = k\sin\theta\cos\phi$, with $k = 2\pi/\lambda$. The range of k_x values when θ,ϕ vary over $0\le\theta\le\pi$ and $0\le\phi\le 2\pi$ is the "visible region". The inversion of the Fourier transform, however, requires knowledge of $F_x(k_x)$ over all

k_x, and in such case the inverse is:

$$I(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F_x(k_x)\,e^{-jk_x x}\,dk_x \tag{20.6.2}$$

Suppose now that the current $I(x)$ is sampled at the regular intervals $x_m = md$ with spacing d and integer m. The sampled current may be represented as the sum of impulses:

$$\hat{I}(x)=\sum_{m=-\infty}^{\infty} I(x_m)\,\delta(x-x_m)=\sum_{m=-\infty}^{\infty} I_m\,\delta(x-md) \tag{20.6.3}$$

where we set $I_m = I(x_m)=I(md)$. Then, the corresponding Fourier transform will be:

$$\hat{F}_x(k_x)=\int_{-\infty}^{\infty}\hat{I}(x)\,e^{jk_x x}dx = \sum_{m=-\infty}^{\infty} I_m\,e^{jmk_x d} = \sum_{m=-\infty}^{\infty} I_m\,e^{jm\psi} \tag{20.6.4}$$

This has precisely the form of an array factor with $\psi = k_x d$. The pattern $\hat{F}_x(k_x)$ is periodic in k_x with period $k_s = 2\pi/d$, which is the sampling frequency in units of radians/meter. Equivalently, $\hat{F}_x(k_x)$ is periodic in ψ with period 2π. The Poisson summation formula [47] relates $\hat{F}_x(k_x)$ to the unsampled pattern $F_x(k_x)$ as a sum of shifted replicas:

$$\hat{F}_x(k_x)=\frac{1}{d}\sum_{n=-\infty}^{\infty} F_x(k_x-nk_s) \tag{20.6.5}$$

Aliasing, that is, the overlapping of the spectral replicas, can be avoided only if $F_x(k_x)$ is bandlimited to within the Nyquist interval, $|k_x|\le k_s/2$. This would imply that $I(x)$ have infinite extent.

In practice, $I(x)$ is assumed to be space-limited with a finite extent, say, over an interval $-l/2\le x\le l/2$. In this case, $F_x(k_x)$ cannot be bandlimited and therefore, aliasing will always occur. However, if the pattern $F(k_x)$ attenuates with large k_x, aliasing may be minimized by selecting a small enough d.

Eqs. (20.6.4) and (20.6.5) provide two equivalent ways to express the spectrum of the sampled current. Eq. (20.6.4) can be inverted to recover the current samples I_m:

$$I_m = \frac{1}{k_s}\int_{-k_s/2}^{k_s/2}\hat{F}_x(k_x)\,e^{-jmk_x d}\,dk_x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{F}_x(\psi)\,e^{-jm\psi}\,d\psi \tag{20.6.6}$$

which is the inverse discrete-space Fourier transform that we introduced in (19.4.8). By using the z-domain variable $z = e^{j\psi}$, (20.6.4) can also be written as the spatial z-transform:

$$\hat{F}_x(z)=\sum_{m=-\infty}^{\infty} I_m\,z^n \tag{20.6.7}$$

Next, we focus on finite line sources $I(x)$, $-l/2\le x\le l/2$. Then, (20.6.1) reads:

$$F_x(k_x)=\int_{-l/2}^{l/2} I(x)\,e^{jk_x x}\,dx \tag{20.6.8}$$

It proves convenient to define a normalized wavenumber variable u by:

$$u=\frac{lk_x}{2\pi}\quad\Leftrightarrow\quad k_x=\frac{2\pi u}{l}\quad\Leftrightarrow\quad u=\frac{l}{\lambda}\sin\theta\cos\phi \tag{20.6.9}$$

and define a scaled pattern $F(u)=F_x(k_x)/l$. Then, we have the Fourier relationships:

$$\boxed{F(u)=\frac{1}{l}\int_{-l/2}^{l/2} I(x)\,e^{j2\pi ux/l}\,dx} \quad\Leftrightarrow\quad \boxed{I(x)=\int_{-\infty}^{\infty}F(u)\,e^{-j2\pi ux/l}\,du} \tag{20.6.10}$$

If $I(x)$ were periodic with period l, then $2\pi/l$ would be its fundamental harmonic and $2\pi u/l$ would be interpreted as the uth harmonic. Indeed, the continuous-line version of the Woodward-Lawson method gives u just such an interpretation. Let us define the periodic extension of the space-limited $I(x)$ with period l to be the sum of its replicas:

$$\tilde{I}(x)=\sum_{n=-\infty}^{\infty} I(x-nl) \tag{20.6.11}$$

Then, $\tilde{I}(x)$, being periodic, could be expanded in a Fourier series with coefficients:

$$\tilde{I}(x)=\sum_{p=-\infty}^{\infty} c_p\,e^{-j2\pi px/l}\,,\quad c_p=\frac{1}{l}\int_{-l/2}^{l/2}\tilde{I}(x)\,e^{j2\pi px/l}\,dx \tag{20.6.12}$$

Because $\tilde{I}(x)=I(x)$ over the period $-l/2\le x\le l/2$, the above integral for the pth coefficient implies from (20.6.10) that $c_p=F(u)$ with $u=p$. Thus, restricting x over its basic period, we have the representation:

$$I(x)=\sum_{p=-\infty}^{\infty}F(p)\,e^{-j2\pi px/l}\,,\quad -\frac{l}{2}\le x\le\frac{l}{2} \tag{20.6.13}$$

The pattern $F(u)$ may itself be expressed in terms of its samples $F(p)$. We have from (20.6.13):

$$F(u)=\frac{1}{l}\int_{-l/2}^{l/2}I(x)\,e^{j2\pi ux/l}\,dx=\sum_{p=-\infty}^{\infty}F(p)\,\frac{1}{l}\int_{-l/2}^{l/2}e^{j2\pi(u-p)x/l}\,dx\,,\quad\text{or,}$$

$$F(u)=\sum_{p=-\infty}^{\infty}F(p)\,\frac{\sin\big(\pi(u-p)\big)}{\pi(u-p)} \tag{20.6.14}$$

Eqs. (20.6.13) and (20.6.14) are the continuous-line version of the Woodward-Lawson method, which is of course equivalent to the application of Shannon's sampling theorem to the space-limited function $I(x)$, and our derivation is nothing more than the proof of that theorem.

For discrete arrays, we must sample in space $x_m=md$, not in frequency. By taking N samples over the length l, that is, $d=l/N$, and truncating the summation in (20.6.13) to $p=0,1,\dots,N-1$, we obtain the practical version of the Woodward-Lawson method that we used in the previous section.

For an N-element finite array, the z-transform $\hat{F}_x(z)$ of Eq. (20.6.7) becomes a polynomial of degree $N-1$ in z. Such an array can be designed directly in discrete-space domain, or it can be designed by mapping a given continuous line source pattern to the discrete case. This can be accomplished approximately by mapping $N-1$ zeros of the continuous pattern to $N-1$ zeros of the array using the mapping $z=e^{j\psi}=e^{jk_xd}$. Since $d=l/N$, the mapping from u-space to ψ-space becomes $\psi=k_xd=2\pi ud/l=2\pi u/N$:

$$\boxed{\psi=k_xd=\frac{2\pi u}{N}} \tag{20.6.15}$$

Therefore, if u_n, $n=1,2,\dots,N-1$ are the $N-1$ zeros of the pattern $F(u)$ on which the design is to be based, then, we may define the corresponding zeros of the array by:

$$\psi_n=\frac{2\pi u_n}{N}\quad\Rightarrow\quad z_n=e^{j\psi_n}=e^{j2\pi u_n/N}\,,\quad n=1,2,\dots,N-1 \tag{20.6.16}$$

and construct the array pattern polynomial from these zeros:

$$A(z)=\prod_{n=1}^{N-1}(z-z_n) \tag{20.6.17}$$

The method is an approximation because $F(u)$ generally has an infinity of zeros. However, good results are obtained if N is large (e.g., $N>10$).

To clarify the above definitions and Fourier relationships, we consider three examples: (a) the uniform line source and how it relates to the uniform array, (b) Taylor's one-parameter line source and its use to design Taylor-Kaiser arrays, and (c) Taylor's ideal line source, which is an idealization of the Chebyshev array, and leads to the so-called Taylor's $\bar{n}$ distribution. A uniform line source has constant current:

$$I(x)=\begin{cases}1\,, & \text{if } -l/2\le x\le l/2\\ 0\,, & \text{otherwise}\end{cases} \tag{20.6.18}$$

Its pattern is:

$$F(u)=\frac{1}{l}\int_{-l/2}^{l/2}I(x)\,e^{j2\pi ux/l}\,dx=\frac{1}{l}\int_{-l/2}^{l/2}e^{j2\pi ux/l}\,dx=\frac{\sin(\pi u)}{\pi u} \tag{20.6.19}$$

Its zeros are at the non-zero integers $u_n=\pm n$, for $n=1,2,\dots$. By selecting the first $N-1$ of these, $u_n=n$, for $n=1,2,\dots,N-1$, we may map them to the $N-1$ zeros of the uniform array:

$$z_n=e^{j2\pi u_n/N}=e^{j2\pi n/N}\,,\quad n=1,2,\dots,N-1$$

The constructed array polynomial will be then,

$$A(z)=\frac{1}{N}\prod_{n=1}^{N-1}(z-z_n)=\frac{1}{N}\prod_{n=1}^{N-1}\big(z-e^{j2\pi n/N}\big)=\frac{1}{N}\,\frac{\displaystyle\prod_{n=0}^{N-1}\big(z-e^{j2\pi n/N}\big)}{z-1}$$

where we introduced a scale factor $1/N$ and multiplied and divided by the factor $(z-1)$. But the numerator polynomial, being a monic polynomial and having as roots the Nth roots of unity, must be equal to z^N-1. Thus,

$$A(z)=\frac{1}{N}\,\frac{z^N-1}{z-1}=\frac{1}{N}\big(1+z+z^2+\cdots+z^{N-1}\big)$$

which has uniform array weights, $a_m = 1/N$. Replacing $z = e^{j\psi} = e^{j2\pi u/N}$, we have:

$$A(\psi) = \frac{1}{N}\,\frac{e^{j\psi N}-1}{e^{j\psi}-1} = \frac{\sin(N\psi/2)}{N\sin(\psi/2)}\,e^{j\psi(N-1)/2} = \frac{\sin(\pi u)}{N\sin(\pi u/N)}\,e^{j\pi u(N-1)/N}$$

For large N and fixed value of u, we may use the approximation $\sin x \simeq x$ in the denominator which tends to $N\sin(\pi u/N)\simeq N(\pi u/N)=\pi u$, thus, approximating the $\sin\pi u/\pi u$ pattern of the continuous line case.

Taylor's one-parameter continuous line source [1085] has current $I(x)$ and corresponding pattern $F(u)$ given by the Fourier transform pair [156]:

$$F(u) = \frac{\sinh\left(\pi\sqrt{B^2-u^2}\right)}{\pi\sqrt{B^2-u^2}} \quad\Leftrightarrow\quad I(x) = I_0\left(\pi B\sqrt{1-(2x/l)^2}\right) \tag{20.6.20}$$

where $-l/2 \le x \le l/2$ and $I_0(\cdot)$ is the modified Bessel function of first kind and zeroth order, and B is a positive parameter that controls the sidelobe level. For $u > B$, the pattern becomes a sinc-pattern in the variable $\sqrt{u^2-B^2}$, and for large u, it tends to the pattern of the uniform line source. We will discuss this further in Sec. 20.10.

Taylor's ideal line source [1086] also has a parameter that controls the sidelobe level and is is defined by the Fourier pair [156]:

$$\begin{aligned} F(u) &= \cosh\left(\pi\sqrt{A^2-u^2}\right)\\ I(x) &= \frac{I_1\left(\pi A\sqrt{1-(2x/l)^2}\right)}{\sqrt{1-(2x/l)^2}}\,\frac{\pi A}{l} + \delta\left(x-\frac{l}{2}\right) + \delta\left(x+\frac{l}{2}\right)\end{aligned} \tag{20.6.21}$$

where $I_1(\cdot)$ is the modified Bessel function of first kind and first order. Van der Maas [1074] showed first that this pair is the limit of a Dolph-Chebyshev array in the limit of a large number of array elements. We will explore it further in Sec. 20.12.

20.7 Narrow-Beam Low-Sidelobe Designs

The problem of designing arrays having narrow beams with low sidelobes is equivalent to the DSP problem of spectral analysis of *windowed sinusoids*. A single beam corresponds to a single sinusoid, multiple beams to multiple sinusoids.

To understand this equivalence, suppose one wants to design an infinitely narrow beam toward some look direction $\phi = \phi_0$. In ψ-space, the array factor (spatial or wavenumber spectrum) should be the infinitely thin spectral line:[†]

$$A(\psi) = 2\pi\delta(\psi-\psi_0)$$

where $\psi = kd\cos\phi$ and $\psi_0 = kd\cos\phi_0$. Inserting this into the inverse DSFT of Eq. (20.3.2), gives the double-sided infinitely-long array, for $-\infty < m < \infty$:

$$a(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi}A(\psi)e^{-jm\psi}d\psi = \frac{1}{2\pi}\int_{-\pi}^{\pi}2\pi\delta(\psi-\psi_0)e^{-jm\psi}d\psi = e^{-j\psi_0 m}$$

[†] To be periodic in ψ, all the Nyquist replicas of this term must be added. But they are not shown here because ψ_0 and ψ are assumed to lie in the central Nyquist interval $[-\pi,\pi]$.

This is the spatial analog of an infinite sinusoid $a(n)=e^{j\omega_0 n}$ whose spectrum is the sharp spectral line $A(\omega)=2\pi\delta(\omega-\omega_0)$. A finite-duration sinusoid is obtained by windowing with a length-N time window $w(n)$ resulting in $a(n)=w(n)e^{j\omega_0 n}$.

In the frequency domain, the effect of windowing is to replace the spectral line $\delta(\omega-\omega_0)$ by its smeared version $W(\omega-\omega_0)$, where $W(\omega)$ is the DTFT of the window $w(n)$. The spectrum $W(\omega-\omega_0)$ exhibits a main lobe at $\omega=\omega_0$ and sidelobes. The main lobe gets narrower with increasing N.

A finite N-element array with a narrow beam and low sidelobes, and steered towards an angle ϕ_0, can be obtained by windowing the infinite narrow-beam array with an appropriate length-N spatial window $w(m)$. For odd $N=2M+1$, or even $N=2M$, we define respectively:

$$\boxed{\begin{aligned} a(m) &= e^{-jm\psi_0}w(m), \quad m = 0,\pm1,\pm2,\dots,\pm M\\ a(\pm m) &= e^{\mp j(m-1/2)\psi_0}w(\pm m), \quad m = 1,2,\dots,M\end{aligned}} \tag{20.7.1}$$

In both cases, the array factor of Eqs. (20.1.1) and (20.1.2) becomes:

$$\boxed{A(\psi) = W(\psi-\psi_0)} \qquad \text{(narrow beam array factor)} \tag{20.7.2}$$

where $W(\psi)$ is the DSFT of the window, defined for odd or even N as:

$$\boxed{\begin{aligned} W(\psi) &= w(0) + \sum_{m=1}^{M}\left[w(m)e^{jm\psi} + w(-m)e^{-jm\psi}\right]\\ W(\psi) &= \sum_{m=1}^{M}\left[w(m)e^{j(m-1/2)\psi} + w(-m)e^{-j(m-1/2)\psi}\right]\end{aligned}} \tag{20.7.3}$$

Assuming a symmetric window, $w(-m)=w(m)$, we can rewrite:

$$\boxed{\begin{aligned} W(\psi) &= w(0) + 2\sum_{m=1}^{M}w(m)\cos(m\psi) && (N=2M+1)\\ W(\psi) &= 2\sum_{m=1}^{M}w(m)\cos\big((m-1/2)\psi\big) && (N=2M)\end{aligned}} \tag{20.7.4}$$

At broadside, $\psi_0=0$, $\phi_0=90^o$, Eq. (20.7.1) reduces to $a(m)=w(m)$ and the array factor becomes $A(\psi)=W(\psi)$. Thus, the weights of a broadside narrow beam array *are* the window samples $a(m)=w(m)$. The steered weights (20.7.1) can be calculated with the help of the MATLAB function `scan`, or `steer`:

```
a = scan(w, psi0);
a = steer(d, w, phi0);
```

The primary issue in choosing a window function $w(m)$ is the tradeoff between frequency resolution and frequency leakage, that is, between *main-lobe width* and *sidelobe level* [47,48]. Ideally, one would like to meet, as best as possible, the two conflicting requirements of having a very narrow mainlobe and very small sidelobes.

Fig. 20.7.1 shows four narrow-beam design examples illustrating this tradeoff. All designs are 7-element arrays with half-wavelength spacing, $d = \lambda/2$, and steered towards 90°. The Dolph-Chebyshev and Taylor-Kaiser arrays were designed with sidelobe level of $R = 20$ dB.

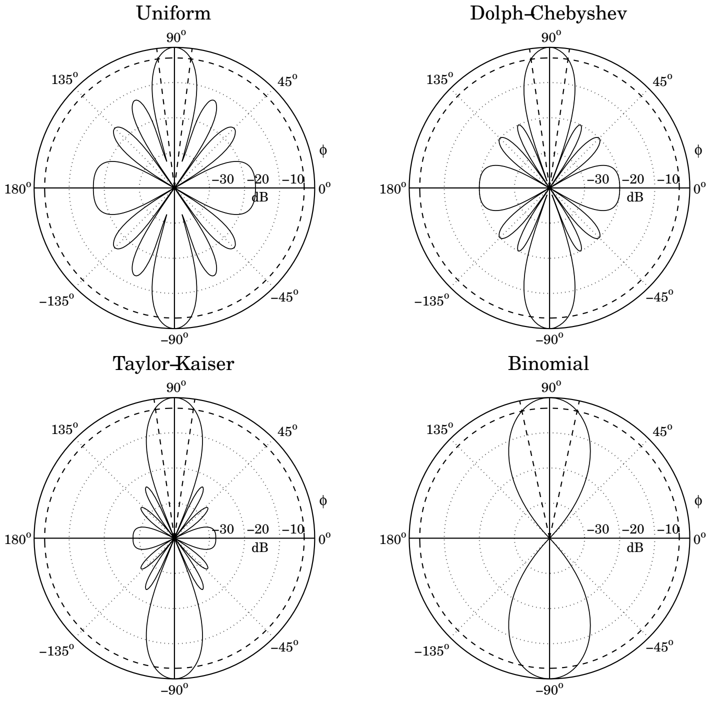


Fig. 20.7.1 Narrow beam design examples.

Shown on the graphs are also the half-power 3-dB circles being intersected by the angular rays at the 3-dB angles. For comparison, we list below the designed array weights (normalized to unity at their endpoints) and the corresponding 3-dB angular widths (in degrees):

Uniform	Dolph-Chebyshev	Taylor-Kaiser	Binomial
1	1.0000	1.0000	1
1	1.2764	1.8998	6
1	1.6837	2.6057	15
1	1.8387	2.8728	20
1	1.6837	2.6057	15
1	1.2764	1.8998	6
1	1.0000	1.0000	1
14.5°	16.4°	16.8°	24.6°

The uniform array has the narrowest mainlobe but also the highest sidelobes. The Dolph-Chebyshev is optimum in the sense that, for the given sidelobe level of 20 dB, it has the narrowest width. The Taylor-Kaiser is somewhat wider than the Dolph-Chebyshev, but it exhibits better sidelobe behavior. The binomial array has the widest mainlobe but no sidelobes at all.

Fig. 20.7.2 shows another set of examples. All designs are 21-element arrays with half-wavelength spacing, $d = \lambda/2$, and scanned towards 60°.

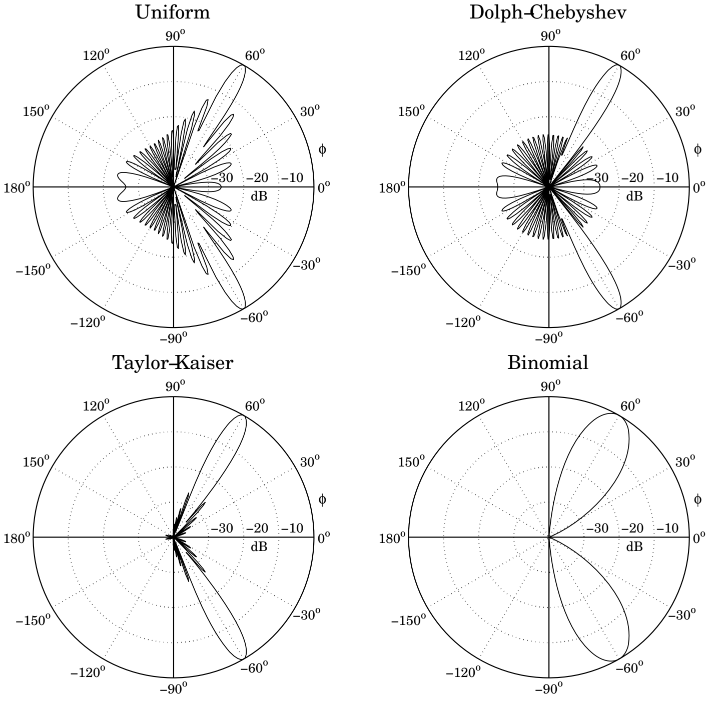


Fig. 20.7.2 Comparison of steered 21-element narrow-beam arrays.

The Dolph-Chebyshev and Taylor arrays were designed with sidelobe level of $R = 25$ dB. The uniform array has sidelobes at $R = 13$ dB. Because N is higher than in Fig. 20.7.1, the beams will be much narrower. The 3-dB beamwidths are in the four cases:

$$\begin{aligned}
\Delta\phi_{3\text{dB}} &= 5.58^\circ && \text{Uniform}\\
\Delta\phi_{3\text{dB}} &= 6.44^\circ && \text{Dolph-Chebyshev}\\
\Delta\phi_{3\text{dB}} &= 7.03^\circ && \text{Taylor-Kaiser}\\
\Delta\phi_{3\text{dB}} &= 15.64^\circ && \text{Binomial}
\end{aligned}$$

The two key parameters characterizing a window are the 3-dB width of its main lobe, $\Delta\psi_{3\text{dB}}$, and its sidelobe level R (in dB). For some windows, such as Dolph-Chebyshev and binomial, $\Delta\psi_{3\text{dB}}$ can be calculated exactly. In others, such as Taylor-Kaiser and Hamming, it can be calculated approximately by Eq. (19.10.2), that is,

$$\boxed{\Delta\psi_{3\text{dB}} = 0.886\,\frac{2\pi b}{N}} \qquad \text{(3-dB width in } \psi\text{-space)} \tag{20.7.5}$$

where b is a broadening factor that depends on the choice of window and increases with the sidelobe attenuation R. As discussed in Sec. 19.10, once $\Delta\psi_{3\text{dB}}$ is known, the angular 3-dB width of the steered array can be computed approximately by:

$$\boxed{\Delta\phi_{3\text{dB}} = \begin{cases} \dfrac{\Delta\psi_{3\text{dB}}}{kd\sin\phi_0}\,, & \text{for } 0^{\text{o}} < \phi_0 < 180^{\text{o}} \\[2ex] 2\sqrt{\dfrac{\Delta\psi_{3\text{dB}}}{kd}}\,, & \text{for } \phi_0 = 0^{\text{o}},\, 180^{\text{o}} \end{cases}} \tag{20.7.6}$$

This is an adequate approximation in practice. In succeeding sections, we discuss the binomial, Dolph-Chebyshev, and Taylor-Kaiser arrays in more detail. In addition, we discuss prolate arrays, Taylor's $\bar{n}$ distribution, and Villeneuve arrays.

We finish this section by summarizing the uniform array, which is based on the rectangular window and has $b = 1$ and sidelobe level $R = 13$ dB. Its weights, symmetric DSFT, and symmetric z-transform were determined in Example 20.1.2:

$$\begin{aligned} \mathbf{w} &= \frac{1}{N}[1, 1, \dots, 1] \\ W(\psi) &= \frac{\sin\left(\dfrac{N\psi}{2}\right)}{N\sin\left(\dfrac{\psi}{2}\right)} \\ W(z) &= \frac{1}{N}\,\frac{z^{N/2} - z^{-N/2}}{z^{1/2} - z^{-1/2}} = z^{-(N-1)/2}\,\frac{1}{N}\,\frac{z^N - 1}{z - 1} \end{aligned} \tag{20.7.7}$$

20.8 Binomial Arrays

The weights of an N-element binomial array are the binomial coefficients:

$$w(m) = \frac{(N-1)!}{m!(N-1-m)!}\,, \qquad m = 0, 1, \dots, N-1 \tag{20.8.1}$$

For example, for $N = 4$ and $N = 5$ they are:

$$\begin{aligned} \mathbf{w} &= [1, 3, 3, 1] \\ \mathbf{w} &= [1, 4, 6, 4, 1] \end{aligned}$$

The binomial weights are the expansion coefficients of the polynomial $(1 + z)^{N-1}$. Indeed, the symmetric z-transform of the binomial array is defined as:

$$W(z) = \left(z^{1/2} + z^{-1/2}\right)^{N-1} = z^{-(N-1)/2}(1 + z)^{N-1} \tag{20.8.2}$$

Setting $z = e^{j\psi}$, we find the array factor in ψ-space:

$$W(\psi) = \left(e^{j\psi/2} + e^{-j\psi/2}\right)^{N-1} = \left[2\cos\left(\frac{\psi}{2}\right)\right]^{N-1} \tag{20.8.3}$$

This response falls monotonically on either side of the peak at $\psi = 0$ until it becomes zero at the Nyquist frequency $\psi = \pm\pi$. Indeed, the z-transform has a multiple zero of order $N - 1$ at $z = -1$.

Thus, the binomial response has no sidelobes. This is, of course, at the expense of a fairly wide mainlobe. The 3-dB width $\Delta\psi_{3\text{dB}}$ can be determined by finding the 3-dB frequencies $\pm\psi_3$ that satisfy the half-power condition:

$$\frac{|W(\psi_3)|^2}{|W(0)|^2} = \frac{1}{2} \qquad \Rightarrow \qquad \left[\cos\left(\frac{\psi_3}{2}\right)\right]^{2(N-1)} = \frac{1}{2}$$

The solution is:

$$\psi_3 = 2\,\text{acos}\left(2^{-0.5/(N-1)}\right)$$

Therefore, the 3-dB width will be $\Delta\psi_{3\text{dB}} = 2\psi_3$:

$$\boxed{\Delta\psi_{3\text{dB}} = 4\,\text{acos}\left(2^{-0.5/(N-1)}\right)} \tag{20.8.4}$$

Once $\Delta\psi_{3\text{dB}}$ is found, the 3-dB width $\Delta\phi_{3\text{dB}}$ in angle space, for an array steered towards an angle ϕ_0, can be found from Eq. (20.7.6). The MATLAB function `binomial` generates the array weights (steered towards ϕ_0) and 3-dB width. Its usage is:

```
[a, dph] = binomial(d, ph0, N);      % binomial array coefficients and beamwidth
```

For example, the fourth graph of the binomial response of Fig. 20.7.1 was generated by the MATLAB code:

```
[a, dph] = binomial(0.5, 90, 5);     % array weights and 3-dB width
[g, ph] = array(0.5, a, 200);        % compute array gain
dbz(ph, g, 45, 40);                  % plot gain in dB with 40-dB scale
addcirc(3, 40, '--');                % add 3-dB grid circle
addray(90 + dph/2, '-');             % add rays at 3-dB angles
addray(90 - dph/2, '-');
```

20.9 Dolph-Chebyshev Arrays

Most windows have largest sidelobes near the main lobe. If a window is designed to achieve a minimum sidelobe attenuation of R dB, then typically R will be the attenuation of the sidelobes nearest to the mainlobe; the sidelobes further away will have attenuations higher than R.

Because of the tradeoff between mainlobe width and sidelobe attenuation, the extra attenuation of the furthest sidelobes will come at the expense of increased mainlobe width. If the attenuation of these sidelobes could be decreased (up to the level of the minimum R), then the mainlobe width would narrow.

It follows that for a given *minimum* desired sidelobe level R, the *narrowest* mainlobe width will be achieved by a window whose sidelobes are all *equal* to R. Conversely, for a given *maximum* desired mainlobe width, the *largest* sidelobe attenuation will be achieved by a window with equal sidelobe levels.

This "optimum" window is the Dolph-Chebyshev window, which is constructed with the help of Chebyshev polynomials. The mth Chebyshev polynomial $T_m(x)$ is:

$$T_m(x) = \cos\big(m\,\mathrm{acos}(x)\big) \qquad (20.9.1)$$

If $|x| > 1$, the inverse cosine $\mathrm{acos}(x)$ becomes imaginary, and the expression can be rewritten in terms of hyperbolic cosines: $T_m(x) = \cosh\big(m\,\mathrm{acosh}(x)\big)$.

Setting $x = \cos\theta$, or $\theta = \mathrm{acos}(x)$, we see that $T_m(x) = \cos(m\theta)$. Using trigonometric identities, the quantity $\cos(m\theta)$ can always be expanded as a polynomial in powers of $\cos\theta$. The expansion coefficients are precisely the coefficients of the powers of x of the Chebyshev polynomial. For example, we have:

$$\begin{array}{lcl}
\cos(0\theta) = 1 & & T_0(x) = 1 \\
\cos(1\theta) = \cos\theta & & T_1(x) = x \\
\cos(2\theta) = 2\cos^2\theta - 1 & \Rightarrow & T_2(x) = 2x^2 - 1 \\
\cos(3\theta) = 4\cos^3\theta - 3\cos\theta & & T_3(x) = 4x^3 - 3x \\
\cos(4\theta) = 8\cos^4\theta - 8\cos^2\theta + 1 & & T_4(x) = 8x^4 - 8x^2 + 1
\end{array}$$

For $|x| < 1$, the Chebyshev polynomial has equal ripples, whereas for $|x| > 1$, it increases like x^m. Moreover, $T_m(x)$ is even in x if m is even, and odd in x if m is odd. Fig. 20.9.1 depicts the Chebyshev polynomials $T_9(x)$ and $T_{10}(x)$.

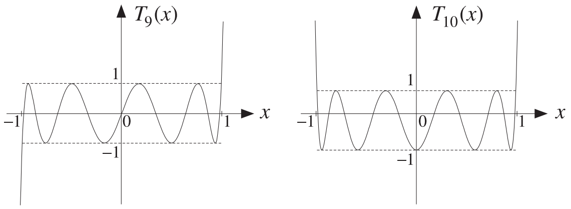


Fig. 20.9.1 Chebyshev polynomials of orders nine and ten.

The Dolph-Chebyshev window is defined such that its sidelobes will correspond to a portion of the equi-ripple range $|x| \le 1$ of the Chebyshev polynomial, whereas its mainlobe will correspond to a portion of the range $x > 1$.

For either even or odd N, Eq. (20.7.4) implies that any window spectrum $W(\psi)$ can be written in general as a polynomial of degree $N-1$ in the variable $u = \cos(\psi/2)$. Indeed, we have for the mth terms:

$$\cos(m\psi) = \cos\left(2m\frac{\psi}{2}\right) = T_{2m}(u)$$

$$\cos\big((m-1/2)\psi\big) = \cos\left((2m-1)\frac{\psi}{2}\right) = T_{2m-1}(u)$$

Thus in the odd case, the summation in Eq. (20.7.4) will result in a polynomial of maximal degree $2M = N-1$ in the variable u, and in the even case, it will result into a polynomial of degree $2M - 1 = N - 1$.

The Dolph-Chebyshev [1068] array factor is defined by the Chebyshev polynomial of degree $N-1$ in the scaled variable $x = x_0\cos(\psi/2)$, that is,

$$\boxed{W(\psi) = T_{N-1}(x)\,, \quad x = x_0\cos\left(\frac{\psi}{2}\right)} \quad \text{(Dolph-Chebyshev array factor)} \qquad (20.9.2)$$

The scale factor x_0 is always $x_0 > 1$ and is determined below. For a broadside design, as the azimuthal angle ϕ ranges over the interval $0^{\circ} \le \phi \le 180^{\circ}$, the wavenumber $\psi = kd\cos\phi$ will range over the visible region $-kd \le \psi \le kd$. The quantity $x = x_0\cos(\psi/2)$ will range from $x_{\min} = x_0\cos(kd/2)$ to the value $x = x_0$, which is reached broadside at $\phi = 90^{\circ}$ or $\psi = 0$, and then x will move back to $x_{\min}$. Thus, the range of variation of x will be $x_{\min} \le x \le x_0$.

Assuming that $x_{\min}$ is in the interval $-1 \le x_{\min} \le 1$, we can split the interval $[x_{\min}, x_0]$ into the two subintervals: $[x_{\min}, 1]$ and $[1, x_0]$, as shown in Fig. 20.9.2. We require that the subinterval $[x_{\min}, 1]$ coincide with the sidelobe interval of the array factor $W(\psi)$, and that the subinterval $[1, x_0]$ coincide with the mainlobe interval. The zeros of the Chebyshev polynomial within $[x_{\min}, 1]$ become the sidelobe zeros of the array factor and get repeated twice as ϕ varies over $[0^{\circ}, 180^{\circ}]$.

In Fig. 20.9.2, for spacing $d = \lambda/2$, we have $kd = \pi$ and $x_{\min} = x_0\cos(kd/2) = x_0\cos(\pi/2) = 0$. Similarly, we have $x_{\min} = x_0\cos(3\pi/4) = -0.707x_0$ for $d = 3\lambda/4$, and $x_{\min} = x_0\cos(\pi/4) = 0.707x_0$ for $d = \lambda/4$.

The relative sidelobe attenuation level in absolute units and in dB is defined in terms of the ratio of the mainlobe to the sidelobe heights:

$$R_a = \frac{W_{\text{main}}}{W_{\text{side}}}\,, \quad R = 20\log_{10}(R_a)\,, \quad R_a = 10^{R/20}$$

Because the mainlobe peak occurs at $\psi = 0$ or $x = x_0$, we will have $W_{\text{main}} = T_{N-1}(x_0)$, and because the sidelobe level is equal to the Chebyshev level within $|x| \le 1$, we will have $W_{\text{side}} = 1$. Thus, we find:

$$\boxed{R_a = T_{N-1}(x_0) = \cosh\big((N-1)\,\mathrm{acosh}(x_0)\big)} \qquad (20.9.3)$$

which can be solved for x_0 in terms of R_a:

$$\boxed{x_0 = \cosh\left(\frac{\mathrm{acosh}(R_a)}{N-1}\right)} \qquad (20.9.4)$$

Once the scale factor x_0 is determined, the window samples $w(m)$ can be computed by constructing the z-transform of the array factor from its zeros and then doing an inverse z-transform. The $N-1$ zeros of $T_{N-1}(x)$ are easily found to be:

$$T_{N-1}(x) = \cos\big((N-1)\,\mathrm{acos}(x)\big) = 0 \quad \Rightarrow \quad x_i = \cos\left(\frac{(i-1/2)\pi}{N-1}\right)$$

for $i = 1, 2, \dots, N-1$. Solving for the corresponding wavenumbers through $x_i = x_0\cos(\psi_i/2)$, we find the pattern zeros:

$$\psi_i = 2\,\mathrm{acos}\Big(\frac{x_i}{x_0}\Big)\,, \qquad z_i = e^{j\psi_i}\,, \qquad i = 1, 2, \dots, N-1$$

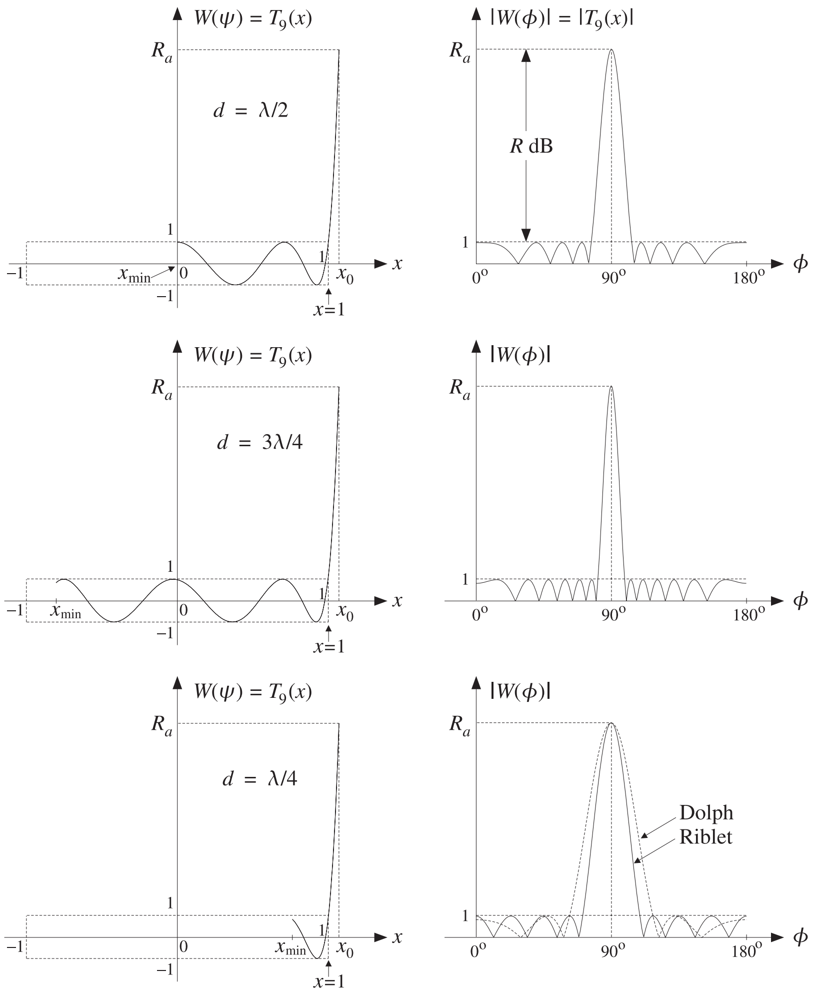


Fig. 20.9.2 Chebyshev polynomials and array factors for $d = \lambda/2$, $d = 3\lambda/4$, and $d = \lambda/4$.

We note that the zeros x_i do not have to lie within the sidelobe range $[x_{\min}, 1]$ and the corresponding ψ_i do not all have to be in the visible region.

The symmetric z-transform of the window is constructed in terms of the one-sided transform using Eq. (20.1.4) as follows:

$$W(z) = z^{-(N-1)/2}\,\tilde{W}(z) = z^{-(N-1)/2}\prod_{i=1}^{N-1}(z - z_i) \tag{20.9.5}$$

The inverse z-transform of $W(z)$ are the window coefficients $w(m)$. The MATLAB

function `dolph.m` of Appendix I implements this design procedure with the help of the function `poly2.m`, which calculates the coefficients from the zeros.[†] The typical MATLAB code in `dolph.m` is as follows:

```
N1 = N-1;                              % number of zeros
Ra = 10^(R/20);                        % sidelobe level in absolute units
x0 = cosh(acosh(Ra)/N1);               % scaling factor

i = 1:N1;
xi = cos(pi*(i-0.5)/N1);               % N1 zeros of Chebyshev polynomial
psi = 2 * acos(xi/x0);                 % N1 array pattern zeros in psi-space
zi = exp(j*psi);                       % N1 zeros of array polynomial

a = real(poly2(zi));                   % zeros-to-polynomial form (N coefficients)
```

The window coefficients resulting from definition (20.9.5) are normalized to unity values at their end-points. This definition differs from that of Eq. (20.9.2) by the scale factor $x_0^{N-1}/2$.

The function `dolph.m` also returns the 3-dB width of the main lobe. The 3-dB frequency ψ_3 is defined by the half-power condition:

$$W(\psi_3) = T_{N-1}(x_3) = \frac{T_{N-1}(x_0)}{\sqrt{2}} = \frac{R_a}{\sqrt{2}} \quad\Rightarrow\quad \cosh\big((N-1)\,\text{acosh}(x_3)\big) = \frac{R_a}{\sqrt{2}}$$

Solving for x_3 and the corresponding 3-dB angle, $x_3 = x_0\cos(\psi_3/2)$, we have:

$$x_3 = \cosh\left(\frac{\text{acosh}(R_a/\sqrt{2})}{N-1}\right), \qquad \psi_3 = 2\,\text{acos}\left(\frac{x_3}{x_0}\right) \tag{20.9.6}$$

which yields the 3-dB width in ψ-space, $\Delta\psi_{3\text{dB}} = 2\psi_3$. The 3-dB width in angle space, $\Delta\phi_{3\text{dB}}$, is then computed from Eq. (20.7.6) or (19.10.6).

There exist several alternative methods for calculating the Chebyshev array coefficients [1072–1080,1082] and have been compared in [1081]. One particularly accurate and effective method is that of Bresler [1075], which has recently been implemented by Simon [1077] with the MATLAB function `chebarray.m`.

Example 20.9.1: The second graph of Fig. 20.7.1 was generated by the MATLAB commands:

```
[a, dph] = dolph(0.5, 90, 5, 20);      % array weights and 3-dB width
[g, ph] = array(0.5, a, 200);          % compute array gain
dbz(ph, g, 45);                        % plot gain in dB
addcirc(3, 40, '--');                  % add 3-dB gain circle
addray(90 + dph/2, '--');              % add 3-dB angles
addray(90 - dph/2, '--');
```

The array weights and 3-dB width were given previously in the table of Fig. 20.7.1. The weights are constructed as follows. The scale parameter x_0 is found to be $x_0 = 1.2933$. The zeros x_i, ψ_i, and z_i are found to be:

i	x_i	ψ_i	z_i
1	0.9239	1.5502	$0.0206 + 0.9998j$
2	0.3827	2.5408	$-0.8249 + 0.5653j$
3	−0.3827	3.7424	$-0.8249 - 0.5653j$
4	−0.9239	4.7330	$0.0206 - 0.9998j$

[†] See Sec. 6.8 regarding the accuracy of `poly2` versus `poly`.

It follows that the one-sided array polynomial will be:

$$\tilde{W}(z) = (z - z_1)(z - z_2)(z - z_3)(z - z_4) = z^4 + 1.6085z^3 + 1.9319z^2 + 1.6085z + 1$$

and the symmetric z-transform:

$$W(z) = z^{-2}\tilde{W}(z) = z^2 + 1.6085z + 1.9319 + 1.6085z^{-1} + z^{-2}$$

resulting in the array weights $\mathbf{w} = [1.0000, 1.6085, 1.9319, 1.6085, 1.0000]$. We note that the array zeros come in conjugate pairs. Only the first two x_i and ψ_i lie in the visible region and show up as pattern zeros in the array factor. □

Example 20.9.2: The second graph of Fig. 20.7.2 was generated by the MATLAB commands:

```
[a, dph] = dolph(0.5, 60, 21, 25);
[g, ph] = array(0.5, a, 200);
dbz(ph, g);
```

The function `dolph.m` was called with the parameters $N = 21$, $R = 20$ dB and was steered towards the angle $\phi_0 = 60^o$. □

Example 20.9.3: As another example, consider the design of a nine-element broadside Dolph-Chebyshev array with half-wavelength spacing and sidelobe attenuation level of $R = 20$ dB. The array factor is shown in Fig. 20.9.2.

The absolute attenuation level is $R_a = 10^{R/20} = 10^{20/20} = 10$, that is, if the peak is normalized to height $R_a = 10$, the sidelobes will have height of unity. The scale factor x_0 is found to be $x_0 = 1.0708$, and the array weights:

$$\mathbf{w} = [1.0000,\ 1.0231,\ 1.3503,\ 1.5800,\ 1.6627,\ 1.5800,\ 1.3503,\ 1.0231,\ 1.0000]$$

The array zeros are constructed as follows:

i	x_i	ψ_i	z_i
1	0.9808	0.8260	$0.6778 + 0.7352j$
2	0.8315	1.3635	$0.2059 + 0.9786j$
3	0.5556	2.0506	$-0.4616 + 0.8871j$
4	0.1951	2.7752	$-0.9336 + 0.3583j$
5	−0.1951	3.5080	$-0.9336 - 0.3583j$
6	−0.5556	4.2326	$-0.4616 - 0.8871j$
7	−0.8315	4.9197	$0.2059 - 0.9786j$
8	−0.9808	5.4572	$0.6778 - 0.7352j$

The 3-dB width is found from Eq. (20.9.6) to be $\Delta\phi_{3\text{dB}} = 12.51^o$. □

In order for the Chebyshev interval $[x_{\min}, 1]$ to be mapped onto the sidelobe region of the array factor, we must require that $x_{\min} \geq -1$.

If $d < \lambda/2$, then this condition is automatically satisfied because $kd < \pi/2$ and $x_{\min} = x_0\cos(kd/2) > 0$. (In this case, we must also demand that $x_{\min} \leq 1$. However, as we discuss below, when $d < \lambda/2$ Dolph's construction is no longer optimal and is replaced by the alternative procedure of Riblet.)

If $\lambda/2 < d < \lambda$, then $\pi < kd < 2\pi$ and $x_{\min} < 0$ and can exceed the left limit $x = -1$. This requires that for the given sidelobe level R, the array spacing may not exceed a *maximum* value that satisfies $x_{\min} = x_0\cos(kd_{\max}/2) = -1$. This gives:

$$kd_{\max} = 2\,\text{acos}\left(-\frac{1}{x_0}\right) \quad\Rightarrow\quad d_{\max} = \frac{\lambda}{\pi}\,\text{acos}\left(-\frac{1}{x_0}\right) \tag{20.9.7}$$

An alternative way of phrasing the condition $x_{\min} \geq -1$ is to say that for the given value of the array spacing d (such that $\lambda/2 < d < \lambda$), there is a *maximum sidelobe attenuation* that may be designed. The corresponding maximum value of x_0 will satisfy $x_{\min} = x_{0,\max}\cos(kd/2) = -1$, which gives:

$$x_{0,\max} = -\frac{1}{\cos(kd/2)} \quad\Rightarrow\quad R_{a,\max} = T_{N-1}(x_{0,\max}) \tag{20.9.8}$$

Example 20.9.4: Consider the case $d = 3\lambda/4$, $R = 20$ dB, $N = 9$. Then for the given R, the maximum element spacing that we can have is $d_{\max} = 0.8836\lambda$.

Alternatively, for the given spacing $d = 3\lambda/4$, the maximum sidelobe attenuation that we can have is $R_{a,\max} = 577$, or, $R_{\max} = 55.22$ dB.

An array designed with the maximum spacing $d = d_{\max}$ will have the narrowest mainlobe, because its total length will be the longest possible. For example, the following two calls to the function `dolph` will calculate the required 3-dB beamwidths:

```
[w, dph1] = dolph(0.75, 90, 9, 20);          % spacing d = 3/4
[w, dph2] = dolph(0.8836, 90, 9, 20);        % spacing d = dmax
```

We find $\Delta\phi_1 = 8.34^o$ and $\Delta\phi_2 = 7.08^o$. The array weights $\mathbf{w}$ are the *same* in the two cases and equal to those of Example 20.9.3. The gains are shown in Fig. 20.9.3. □

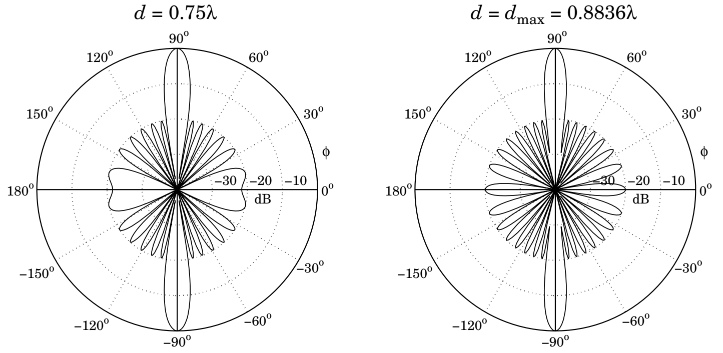


Fig. 20.9.3 Chebyshev arrays with $N = 9$, $R = 20$ dB, $d = 3\lambda/4$ and $d = 0.8836\lambda$.

As pointed out by Riblet [1069], Dolph's procedure is optimal only for element spacings that are greater than half a wavelength, $d \geq \lambda/2$. For $d < \lambda/2$, it is possible to find another set of window coefficients that would result into a narrower main lobe.

Riblet modified Dolph's method to obtain an optimal design for both cases, $d < \lambda/2$ and $d \geq \lambda/2$, but only for an *odd* number of array elements, $N = 2M + 1$.

It follows from Eq. (20.7.4) that if N is odd, the array factor $W(\psi)$ can be expressed either as a polynomial in the variable $\cos(\psi/2)$ or as a polynomial in the variable $\cos\psi$.

Dolph's original definition of Eq. (20.9.2) used a Chebyshev polynomial $T_{2M}(x)$ of order $2M = N - 1$ in the variable $x = x_0\cos(\psi/2)$. Riblet used instead a Chebyshev polynomial $T_M(y)$ of order M in the new variable $y = A\cos\psi + B$, where the constants A, B are to be determined from the desired spacing d and sidelobe attenuation R. The array factor is defined as:

$$\boxed{W(\psi) = T_M(y), \quad y = A\cos\psi + B} \qquad \text{(Riblet's modification)} \tag{20.9.9}$$

The mainlobe peak of height R_a at $\phi = 90^o$ (or $\psi = 0$) will correspond to a value y_0 such that:

$$\boxed{R_a = T_M(y_0) = \cosh\big(M\,\mathrm{acosh}(y_0)\big)} \tag{20.9.10}$$

which may be solved for y_0:

$$y_0 = \cosh\left(\frac{\mathrm{acosh}(R_a)}{M}\right) \tag{20.9.11}$$

We note that y_0 is related to x_0 of Eq. (20.9.3) by $y_0 = 2x_0^2 - 1$. This follows from the general property of Chebyshev polynomials that:

$$y = 2x^2 - 1 \quad\Rightarrow\quad T_{2M}(x) = T_M(y) \tag{20.9.12}$$

Indeed, setting $x = \cos\theta$ and $y = \cos(2\theta) = 2\cos^2\theta - 1 = 2x^2 - 1$, we have $\theta = \mathrm{acos}(x)$ and $2\theta = \mathrm{acos}(y)$, and therefore:

$$T_{2M}(x) = \cos\big((2M)\theta\big) = \cos\big(M(2\theta)\big) = T_M(y)$$

As the azimuthal angle ϕ varies over $0^o \leq \phi \leq 180^o$ and the wavenumber ψ over the visible region $-kd \leq \psi \leq kd$, the quantity $c = \cos\psi$ will vary from $c = \cos(kd)$ at $\phi = 0^o$ to $c = 1$ at $\phi = 90^o$, and then back to $c = \cos(kd)$ at $\phi = 180^o$.

If $\lambda/2 \leq d \leq \lambda$, then $\pi \leq kd \leq 2\pi$ and $\psi = kd\cos\phi$ will pass through the value $\psi = \pi$ before it reaches the value $\psi = kd$. It follows that the quantity c will go through $c = -1$ before it reaches $c = \cos(kd)$. Thus, in this case the widest range of variation of $c = \cos\psi$ is $-1 \leq c \leq 1$.

On the other hand, if $d < \lambda/2$, then $kd < \pi$ and c never reaches the value $c = -1$. Its minimum value is $c = \cos(kd)$, and the range of c is $[\cos(kd), 1]$. To summarize, the range of variation of c will be the interval $[c_0, 1]$, where

$$\boxed{c_0 = \begin{cases} -1, & \text{if } d \geq \lambda/2 \\ \cos(kd), & \text{if } d < \lambda/2 \end{cases}} \tag{20.9.13}$$

Assuming $A > 0$, it follows that the range of variation of $y = A\cos\psi + B$ will be the interval $[Ac_0 + B,\ A + B]$. The parameters A, B are fixed by requiring that this interval coincide with the interval $[-1, y_0]$ so that the right end will correspond to the mainlobe peak, while the left end will ensure that we use the maximum size of the equi-ripple interval of the Chebyshev variable y. Thus, we require the conditions:

$$\begin{aligned} Ac_0 + B &= -1 \\ A + B &= y_0 \end{aligned} \tag{20.9.14}$$

which may be solved for A, B:

$$\begin{aligned} A &= \frac{1 + y_0}{1 - c_0} \\ B &= -\frac{1 + y_0 c_0}{1 - c_0} \end{aligned} \tag{20.9.15}$$

For $d \geq \lambda/2$, the method coincides with Dolph's original method. In this case, $c_0 = -1$, and A, B become:

$$\begin{aligned} A &= \frac{y_0 + 1}{2} = x_0^2 \\ B &= \frac{y_0 - 1}{2} = x_0^2 - 1 \end{aligned} \tag{20.9.16}$$

where we used $y_0 = 2x_0^2 - 1$, as discussed above. It follows that the y variable will be related to the Dolph variable $x = x_0\cos(\psi/2)$ by:

$$y = x_0^2\cos\psi + x_0^2 - 1 = x_0^2(\cos\psi + 1) - 1 = 2x_0^2\cos^2\left(\frac{\psi}{2}\right) - 1 = 2x^2 - 1$$

and therefore, Eq. (20.9.12) implies that $W(\psi) = T_M(y) = T_{2M}(x)$.

Once the parameters A, B are determined, the window $w(m)$ may be constructed from the zeros of the Chebyshev polynomials. The M zeros of $T_M(y)$ are:

$$y_i = \cos\left(\frac{(i - 1/2)\pi}{M}\right), \quad i = 1, 2, \dots, M$$

The corresponding wavenumbers are found by inverting $y_i = A\cos\psi_i + B$:

$$\psi_i = \mathrm{acos}\left(\frac{y_i - B}{A}\right), \quad i = 1, 2, \dots, M$$

The $2M = N - 1$ zeros of the z-transform of the array are the conjugate pairs:

$$\{e^{j\psi_i}\ e^{-j\psi_i}\}, \quad i = 1, 2, \dots, M$$

The symmetrized z-transform will be then:

$$W(z) = z^{-M}\tilde{W}(z) = z^{-M}\prod_{i=1}^{M}\big((z - e^{j\psi_i})(z - e^{-j\psi_i})\big)$$

The inverse z-transform of $W(z)$ will be the desired array weights $w(m)$. This procedure is implemented by the MATLAB function `dolph2.m` of Appendix I. We note again that this definition differs from that of Eq. (20.9.9) by the scale factor $A^M/2$.

The function `dolph2` also returns the 3-dB width of the main lobe. The 3-dB frequency ψ_3 is computed from the half-power condition:

$$W(\psi_3) = T_M(y_3) = \frac{T_M(y_0)}{\sqrt{2}} = \frac{R_a}{\sqrt{2}} \quad \Rightarrow \quad y_3 = \cosh\left(\frac{\text{acosh}(R_a/\sqrt{2})}{M}\right)$$

Inverting $y_3 = A\cos\psi_3 + B$, we obtain the 3-dB width in ψ-space:

$$\psi_3 = \text{acos}\left(\frac{y_3 - B}{A}\right), \qquad \Delta\psi_{\text{3dB}} = 2\psi_3 \tag{20.9.17}$$

For the case $d \geq \lambda/2$, the maximum element spacing given by Eq. (20.9.7) can also be expressed in terms of the variable y_0 as follows:

$$d_{\max} = \lambda\left[1 - \frac{1}{2\pi}\,\text{acos}\left(\frac{3 - y_0}{1 + y_0}\right)\right] \tag{20.9.18}$$

This follows from the condition $x_{\min} = x_0\cos(kd_{\max}/2) = -1$. The corresponding value of y will be $y = 2x_{\min}^2 - 1 = 1$. Using Eq. (20.9.16), this condition reads:

$$y = \frac{y_0 + 1}{2}\cos(kd_{\max}) + \frac{y_0 - 1}{2} = 1 \quad \Rightarrow \quad \cos(kd_{\max}) = \frac{3 - y_0}{1 + y_0}$$

Because the function acos always returns a value in the range $[0, \pi]$, and we want a value $kd_{\max} > \pi$, we must invert the cosine as follows:

$$kd_{\max} = 2\pi - \text{acos}\left(\frac{3 - y_0}{1 + y_0}\right)$$

which implies Eq. (20.9.18).

Example 20.9.5: The bottom two graphs of Fig. 20.9.2 show the array factor designed using Dolph's and Riblet's methods for the case $N = 9$, $R = 20$ dB, and $d = \lambda/4$. The Dolph weights are the same as those given in Example 20.9.3. The Riblet weights computed by `dolph2` are:

$$\mathbf{w} = [1, -3.4884, 7.8029, -11.7919, 13.6780, -11.7919, 7.8029, -3.4884, 1]$$

The corresponding array gains in dB are shown in Fig. 20.9.4. The 3-dB widths of the Dolph and Riblet designs are $\Delta\phi_{\text{3dB}} = 25.01^{\text{o}}$ and $\Delta\phi_{\text{3dB}} = 17.64^{\text{o}}$. ☐

Next, we discuss steered arrays [1070]. We assume a steering angle $0^{\text{o}} < \phi_0 < 180^{\text{o}}$. The endfire case $\phi_0 = 0^{\text{o}}, 180^{\text{o}}$ will be treated separately [1071]. The steered wavenumber will be:

$$\psi' = \psi - \psi_0 = kd(\cos\phi - \cos\phi_0) \tag{20.9.19}$$

where $\psi_0 = kd\cos\phi_0$. The corresponding array weights and array factor will be:

$$\begin{aligned} a(m) &= e^{-jm\psi_0}w(m)\,, \quad -M \leq m \leq M \\ A(\psi) &= W(\psi - \psi_0) = W(\psi') = T_M(y')\,, \qquad y' = A\cos\psi' + B \end{aligned} \tag{20.9.20}$$

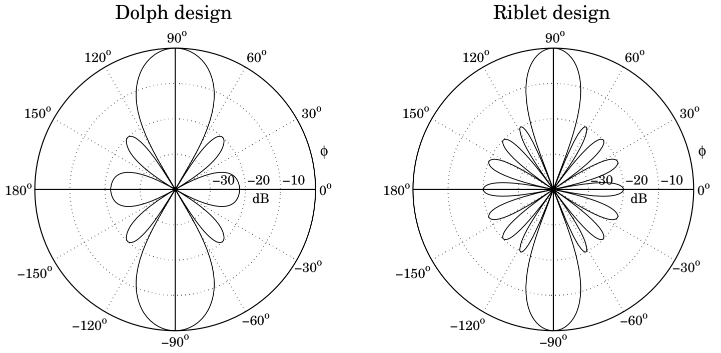


Fig. 20.9.4 Dolph and Riblet designs of Chebyshev array with $N = 9$, $R = 20$ dB, $d = \lambda/4$.

where we assumed that N is odd, $N = 2M + 1$. The visible region becomes now:

$$kd(1 - |\cos\phi_0|) \leq \psi' \leq kd(1 + |\cos\phi_0|)$$

In order to avoid grating lobes, the element spacing must be less than the maximum:

$$\boxed{d_0 = \frac{\lambda}{1 + |\cos\phi_0|}} \tag{20.9.21}$$

which satisfies $kd_0(1 + |\cos\phi_0|) = 2\pi$.

The Chebyshev design method is carried out in the same way, except instead of using the half-wavelength spacing $\lambda/2$ as the dividing line between the Riblet and the Dolph methods, we must use $d_0/2$. Thus, the variable $c = \cos\psi' = \cos(\psi - \psi_0)$ will vary in the interval $[c_0, 1]$, where Eq. (20.9.13) is now replaced by

$$c_0 = \begin{cases} -1, & \text{if } d \geq d_0/2 \\ \cos\bigl(kd(1 + |\cos\phi_0|)\bigr), & \text{if } d < d_0/2 \end{cases}$$

Replacing $1 + |\cos\phi_0| = \lambda/d_0$, we can rewrite this as follows:

$$\boxed{c_0 = \begin{cases} -1, & \text{if } d \geq d_0/2 \\ \cos\left(\dfrac{2\pi d}{d_0}\right), & \text{if } d < d_0/2 \end{cases}} \tag{20.9.22}$$

The solutions for A, B will still be given by Eq. (20.9.15) with this new value for c_0. Note that when $d < d_0/2$ the quantities A, B, and hence the array weights $w(m)$, will depend on ϕ_0. Therefore, the weights must be redesigned for each new value of ϕ_0, instead of simply steering the broadside weights [1070].

When $d \geq d_0/2$, we have $c_0 = -1$ and the weights $w(m)$ become *independent* of ϕ_0. In this case, the steered weights are obtained by steering the broadside weights.

Example 20.9.6: Fig. 20.9.5 shows the gain of an array steered towards $\phi_0 = 60^{\text{o}}$, with $N = 9$, $R = 20$ dB, and element spacing $d = \lambda/4$.

The grating lobe spacing is $d_0 = \lambda/(1+\cos(60^o)) = 2\lambda/3$, and the dividing line between Dolph and Riblet designs will be $d_0/2 = \lambda/3$. The second graph shows the gain of a broadside array, which is steered towards 60°. It demonstrates that the plain steering of a broadside design will not work for $d < d_0/2$. The array weights were computed by the MATLAB commands:

```
a1 = dolph2(1/4, 60, 9, 20);          % steered array
w  = dolph2(1/4, 90, 9, 20);          % broadside array
a2 = steer(1/4, w, 60);               % steered broadside array
```

The 3-dB width was $\Delta\phi_{3\text{dB}} = 26.66^o$. It was obtained using Eq. (20.9.17) and the approximation Eq. (20.7.6). The first graph also shows the 3-dB gain circle intersecting the rays at the 3-dB angles $\phi_0 \pm \Delta\phi_{3\text{dB}}/2$, that is, at 46.67° and 73.33°. We note also that the broadside weights **w** were given in Example 20.9.5. □

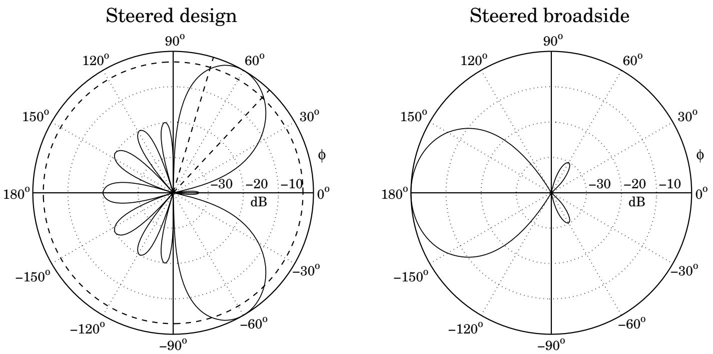


Fig. 20.9.5 Nine-element array with $d = \lambda/4$ steered towards 60°.

Endfire Dolph-Chebyshev arrays require special treatment. DuHamel has shown how to modify Riblet's design for this purpose [1071]. The key idea is not to use a steering angle $\phi_0 = 0^o$ or $\phi_0 = 180^o$, but rather to make ϕ_0, and the corresponding steering phase $\psi_0 = kd\cos\phi_0$, a free design parameter.

The steered wavenumber will still be $\psi' = kd\cos\phi - kd\cos\phi_0 = kd\cos\phi - \psi_0$ and the array factor and array weights will still be given by Eq. (20.9.20).

The three parameters $\{A, B, \psi_0\}$ are determined by the following conditions. For a forward endfire array (with mainlobe peak towards $\phi = 0^o$,) we require that $y' = y_0$ at $\phi = 0$, or, at $\psi' = kd - \psi_0$. Moreover, we require that the two endpoints $y' = -1$ and $y' = 1$ of the equi-ripple range of the Chebyshev polynomial are reached at $\psi' = 0$ and at $\phi = 180^o$, or, $\psi' = -kd - \psi_0$. These three conditions can be stated as follows:

$$\begin{aligned} A\cos(kd-\psi_0)+B &= y_0 \\ A+B &= -1 \\ A\cos(kd+\psi_0)+B &= 1 \end{aligned} \tag{20.9.23}$$

For a backward endfire array (with mainlobe towards $\phi = 180^o$,) we must replace ψ_0 by $-\psi_0$. The solution of Eqs. (20.9.23) is:

$$\boxed{\begin{aligned} A &= -\frac{y_0+3+2\cos(kd)\sqrt{2(y_0+1)}}{2\sin^2(kd)} \\ B &= -1-A \\ \psi_0 &= \pm\,\text{asin}\left(\frac{y_0-1}{2A\sin(kd)}\right) \end{aligned}} \tag{20.9.24}$$

where in the solution for ψ_0, the plus (minus) sign is chosen for the forward (backward) endfire array. Bidirectional endfire arrays can also be designed. In that case, we set $\psi_0 = 0$ and only require the first two conditions in (20.9.23), which become

$$\begin{aligned} A\cos(kd)+B &= y_0 \\ A+B &= -1 \end{aligned} \tag{20.9.25}$$

with solution:

$$\boxed{\begin{aligned} A &= -\frac{y_0+1}{1-\cos(kd)} \\ B &= \frac{y_0+\cos(kd)}{1-\cos(kd)} \end{aligned}} \tag{20.9.26}$$

In all three of the above endfire designs, we must assume $d \le \lambda/2$ in order to avoid grating lobes. The MATLAB function `dolph3.m` of Appendix I implements all three cases.

Example 20.9.7: Fig. 20.9.6 shows three endfire designs for a nine-element array with quarter-wavelength spacing $d = \lambda/4$, and sidelobe level of $R = 20$ dB. The array weights and 3-dB widths were computed as follows:

```
[a1, dph1] = dolph3(1,  1/4, 9, 20);       % forward endfire
[a2, dph2] = dolph3(-1, 1/4, 9, 20);       % backward endfire
[a3, dph3] = dolph3(2,  1/4, 9, 20);       % bidirectional endfire
```

The first argument of `dolph3` takes on one of the three values $\{1, -1, 2\}$, for forward, backward, and bidirectional designs. In the forward and backward cases, the array weights are already scanned by the effective scanning phase $\pm\psi_0$. The calculated array weights are in the three cases:

weights	forward	backward	bidirectional
a_0	18.3655	18.3655	20.4676
$a_1 = a_{-1}^*$	$-15.8051 - 1.0822j$	$-15.8051 + 1.0822j$	-17.5583
$a_2 = a_{-2}^*$	$9.8866 + 1.3603j$	$9.8866 - 1.3603j$	10.8723
$a_3 = a_{-3}^*$	$-4.1837 - 0.8703j$	$-4.1837 + 0.8703j$	-4.5116
$a_4 = a_{-4}^*$	$0.9628 + 0.2701j$	$0.9628 - 0.2701j$	1.0000

Because the backward case is obtained by the replacement $\psi_0 \to -\psi_0$, its weights will be the conjugates of those of the forward case.

The 3-dB widths are in the three cases: $\Delta\phi_{3\text{dB}} = 22.85^o, 22.85^o, 22.09^o$. The graphs also show the 3-dB gain circles intersecting the gains at the 3-dB angles. □

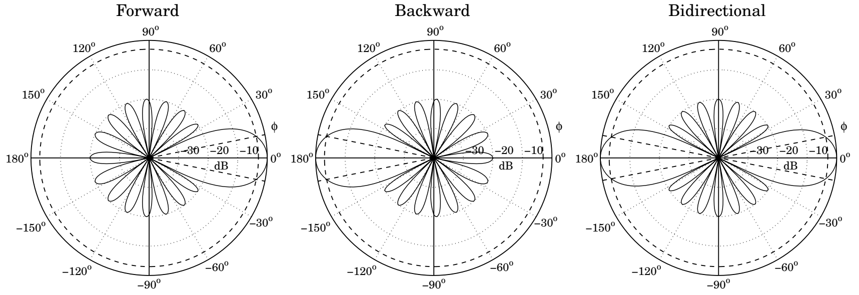

Forward Backward Bidirectional

Fig. 20.9.6 Forward, backward, and bidirectional endfire designs.

20.10 Taylor One-Parameter Source

In Sec. 20.4, we used the Kaiser window to design a sector array pattern. That design problem was equivalent to designing an FIR lowpass digital filter using the window method. Here, we use the Kaiser window to design a narrow beam array—a problem equivalent to the spectral analysis of windowed sinusoids [47,48,1090].

The broadside array weights are equal to the window coefficients $a(m)= w(m)$, defined up to an overall normalization constant by:

$$\boxed{w(m)= I_0\Big(\alpha\sqrt{1-m^2/M^2}\Big)} \tag{20.10.1}$$

where $m = \pm1, \pm2, \dots, \pm M$, or $m = 0, \pm1, \pm2, \dots, \pm M$, for even or odd number of array elements, $N = 2M$ or $N = 2M + 1$.

This window is based on Taylor's one-parameter continuous line source [1085], and is obtained by setting $x_m = md$ with $d = l/(2M)$ in Eq. (20.6.20), so that $2x_m/l = m/M$,

$$I(x_m)= I_0\left(\pi B\sqrt{1-(2x_m/l)^2}\right) = I_0\left(\pi B\sqrt{1-(m/M)^2}\right)$$

Thus, we note that the Kaiser window shape parameter α is related to Taylor's parameter B by $\alpha = \pi B$. The parameter B or α control the sidelobe level. The continuous line pattern of (20.6.20),

$$F(u)= \frac{\sinh\left(\pi\sqrt{B^2-u^2}\right)}{\pi\sqrt{B^2-u^2}} = \frac{\sin\left(\pi\sqrt{u^2-B^2}\right)}{\pi\sqrt{u^2-B^2}} \tag{20.10.2}$$

has a first null at $u_0 = \sqrt{B^2+1}$, and therefore, the first sidelobe will occur for $u > u_0$. For this range, we must use the sinc-form of $F(u)$ and to find the maximum sidelobe level, we must find the maximum of the sinc function (for argument other than zero). This can be determined, for example, by the MATLAB command:[†]

```
x0=fminbnd('sinc(x)', 1,2, optimset('TolX',eps)); r0 = abs(sinc(x0));
```

[†] MATLAB's sinc function is defined as $\mathrm{sinc}(x)= \sin\pi x/\pi x$.

which yields the values:

$$\begin{aligned} x_0 &= 1.4302966532 \\ r_0 &= \left|\mathrm{sinc}(x_0)\right| = 0.2172336282 \\ R_0 &= -20\log_{10}(r_0)= 13.2614588840\ \text{dB} \end{aligned} \tag{20.10.3}$$

The sidelobe level R_a (in absolute units) is defined as the ratio of the pattern at $u = 0$ to the maximum sidelobe level r_0, that is,

$$R_a = \frac{1}{r_0}\,\frac{\sinh(\pi B)}{\pi B} \tag{20.10.4}$$

and in dB, $R = 20\log_{10}(R_a)$,

$$R = R_0 + 20\log_{10}\left(\frac{\sinh(\pi B)}{\pi B}\right) \tag{20.10.5}$$

To avoid having to solve (20.10.4) for B for a given R_a, Kaiser and Schafer [1090] have developed an empirical formula in terms of the sidelobe level R in dB, which is valid across the range $13 < R < 120$ dB:

$$\pi B = \begin{cases} 0, & R \le 13.26 \\ 0.76609(R-13.26)^{0.4}+0.09834(R-13.26), & 13.26 < R \le 60 \\ 0.12438(R+6.3), & 60 < R < 120 \end{cases} \tag{20.10.6}$$

For $R \le 13.26$, $w(m)$ becomes the rectangular window. The broadening factor b, and the 3-dB width in ψ-space can also be expressed in terms of the dB sidelobe level R by the following empirical formula valid for $20 < R < 100$ dB:

$$b = 0.01330\,R + 0.9761\,, \qquad \Delta\psi_{3\mathrm{dB}} = 0.886\,\frac{2\pi b}{N} \tag{20.10.7}$$

The 3-dB width in angle space, $\Delta\phi_{3\mathrm{dB}}$, is then calculated from Eq. (20.7.6). The 3-dB beam width may be more accurately calculated by finding it in u-space, say Δu, and then transforming it to ψ-space using Eq. (20.6.15), $\Delta\psi_{3\mathrm{dB}} = 2\pi\Delta u/N$. The width Δu is given by $\Delta u = 2u_3$, where u_3 is the solution of the half-power condition:

$$\left|F(u_3)\right|^2 = \frac{1}{2}\left|F(0)\right|^2 \quad\Rightarrow\quad \frac{\sinh\left(\pi\sqrt{B^2-u_3^2}\right)}{\pi\sqrt{B^2-u_3^2}} = \frac{1}{\sqrt{2}}\,\frac{\sinh(\pi B)}{\pi B} \tag{20.10.8}$$

For small values of B, the right-hand side becomes less than one, and we must switch the left-hand side to its sinc form. This happens when $B \le B_c$, where

$$\frac{1}{\sqrt{2}}\,\frac{\sinh(\pi B_c)}{\pi B_c} = 1 \quad\Rightarrow\quad B_c = 0.4747380492 \tag{20.10.9}$$

which, through (20.10.5), corresponds to a sidelobe attenuation of $R_c = 16.27$ dB. Rather than using the above empirical formulas, Eqs. (20.10.4) and (20.10.8) may be solved numerically in MATLAB. The function `taylorbw` implements the solution, returning the values of B and Δu, for any vector of sidelobe attenuations R:

```
[B,Du] = taylorbw(R);     % Taylor parameter B and beamwidth Δu
```

It is built on the functions `sinhc` and `asinhc` for computing the hyperbolic sinc function and its inverse:

```
y = sinhc(x);      % hyperbolic sinc function, sinhc(x) = sinh(π x)/π x
x = asinhc(y);     % inverse function, finds the x that satisfies sinhc(x) = y
```

For small x, the equation $y = \sinh(x)/x$ is solved for x by using the Taylor series expansion $y = \sinh(x)x \simeq 1 + x^2/6 + x^4/120$; for larger x, it is solved by the iteration $\sinh(x_n)/x_{n-1} = y$, or, $x_n = \text{asinh}(yx_{n-1})$, for $n = 1, 2, \dots$.

Once the B-parameter is determined, the array weights $w(m)$ can be computed from (20.10.1) using the built-in function `besseli`, and then steered towards an angle ϕ_0 using Eq. (20.7.1). In this case, to avoid grating lobes, the element spacing must be less than the maximum:

$$\boxed{d_0 = \frac{\lambda}{1 + |\cos\phi_0|}} \tag{20.10.10}$$

As discussed in Sec. 20.9, in order for the visible region is ψ-space to cover at least one Nyquist period, the element spacing d must be in the range:

$$\frac{d_0}{2} \le d < d_0 \tag{20.10.11}$$

The MATLAB function `taylor1p` of Appendix I implements this design procedure and invokes the function `taylorbw`. The outputs of the function are the steered array weights and the 3-dB width. It has usage:

```
[a, dph] = taylor1p(d, ph0, N, R);     % Taylor 1-parameter line source
```

Example 20.10.1: Fig. 20.10.1 depicts the gain of a 14- and a 15-element Taylor-Kaiser array with half-wavelength spacing $d = \lambda/2$, steered towards $\phi_0 = 60^o$. The sidelobe level was $R = 20$ dB. The array weights were obtained by:

```
[a1, dph1] = taylor1p(0.5, 60, 14, 20);
[a2, dph2] = taylor1p(0.5, 60, 15, 20);
```

The graphs in Fig. 20.10.1 can be produced by the following commands:

```
[g1,ph1] = gain1d(0.5, a1, 720);             % compute normalized gain at 720 angles
dbz(ph1,g1);                                 % make azimuthal plot of the gain
addcirc(3);                                  % add 3-dB circle
addray(60-dph1/2); addray(60+dph1/2);        % add rays at 3-dB angles
```

The array weights are already steered towards 60°. The designed unsteered weights were in the two cases:

$$\begin{aligned}\mathbf{w}_1 = [&1.0000, 1.3903, 1.7762, 2.1339, 2.4401, 2.6749, 2.8224,\\ &2.8224, 2.6749, 2.4401, 2.1339, 1.7762, 1.3903, 1.0000]\end{aligned}$$

$$\begin{aligned}\mathbf{w}_2 = [&1.0000, 1.3903, 1.7762, 2.1339, 2.4401, 2.6749, 2.8224, 2.8728\\ &2.8224, 2.6749, 2.4401, 2.1339, 1.7762, 1.3903, 1.0000]\end{aligned}$$

The corresponding 3-dB widths were $\Delta\phi_{3\text{dB}} = 9.68^o$ and $\Delta\phi_{3\text{dB}} = 9.03^o$, with the second being slightly narrower because the array is slightly longer. The graphs show how the rays at the two 3-dB angles intersect the 3-dB gain circles. The maximum and minimum array spacings are from (20.10.10): $d_0 = 2\lambda/3$ and $d_0/2 = \lambda/3$. □

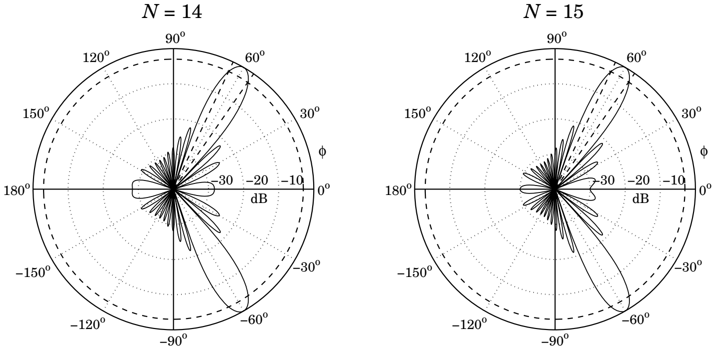


Fig. 20.10.1 Taylor-Kaiser arrays with $N = 14$ and $N = 15$, and $d = \lambda/2$.

Example 20.10.2: Fig. 20.10.2 depicts the gain of a 31-element endfire array with spacing $d = \lambda/4$ and sidelobe level $R = 20$ dB, steered towards the forward direction, $\phi_0 = 0^o$, and the backward one, $\phi_0 = 180^o$.

The maximum and minimum array spacings, calculated from Eq. (20.10.10) for $\phi_0 = 0^o$ and $\phi_0 = 180^o$, are $d_0 = \lambda/2$ and $d_0/2 = \lambda/4$. We have chosen $d = d_0/2 = \lambda/4$.

The 3-dB widths are in both cases $\Delta\phi_{3\text{dB}} = 43.12^o$. The graphs also show the 3-dB circle intersecting the 3-dB angle rays. □

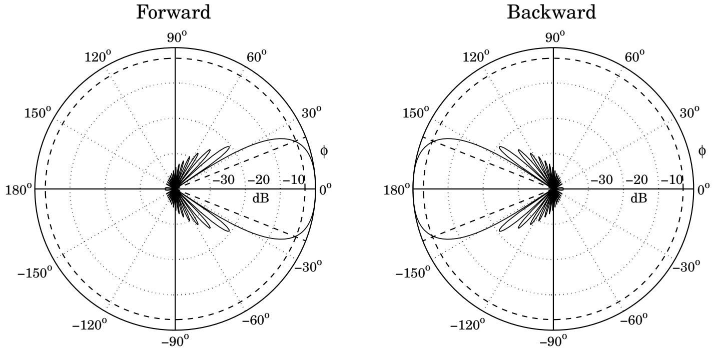


Fig. 20.10.2 Taylor-Kaiser endfire arrays with $N = 31$ and $d = \lambda/4$.

The design method of the section was based on sampling the current distribution directly, as in Eq. (20.10.1), rather than using the procedure of mapping the pattern zeros as outlined in Eqs. (20.6.15)–(20.6.17). A variation of the design method array that uses the latter procedure is discussed in Problem 20.1.

20.11 Prolate Array

Kaiser has noted [1089,1090] that the Kaiser window function (20.10.1) is an excellent approximation to the 0th order discrete prolate spheroidal sequence that maximizes the energy concentration in a given frequency interval [1092–1101].

Using the prolate sequence as a window for array design provides a slight improvement over the Taylor-Kaiser case in the sense of having a slightly narrower beamwidth while meeting the sidelobe specification more precisely. The prolate array can be designed very quickly using the inverse power iteration.

Given an N-dimensional array $\mathbf{a} = [a_0, a_1, \dots, a_{N-1}]^T$ with array pattern $A(\psi)$, the discrete prolate spheroidal performance index which measures the concentration of energy within the wavenumber interval $[-\psi_c, \psi_c]$ is defined by:

$$\mathcal{J} = \frac{\dfrac{1}{2\pi}\displaystyle\int_{-\psi_c}^{\psi_c} |A(\psi)|^2\, d\psi}{\dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} |A(\psi)|^2\, d\psi} \tag{20.11.1}$$

The integration range in the denominator may be changed to be the visible region $[-kd, kd]$ if so desired [1094], but the design technique remains essentially the same. Inserting the array pattern,

$$A(\psi) = \sum_{n=0}^{N-1} a_n e^{j\psi n} \tag{20.11.2}$$

into (20.11.1), we may express the performance index $\mathcal{J}$ as a Rayleigh quotient involving the so-called *prolate matrix* [1093,1101]:

$$\mathcal{J} = \frac{\mathbf{a}^\dagger A \mathbf{a}}{\mathbf{a}^\dagger \mathbf{a}} \tag{20.11.3}$$

where the dagger denotes the Hermitian conjugate and the prolate matrix is defined by its matrix elements:

$$A_{nm} = \frac{\sin(\psi_c(n-m))}{\pi(n-m)} = \frac{\sin(2\pi W(n-m))}{\pi(n-m)}\,, \quad n, m = 0, 1, \dots, N-1 \tag{20.11.4}$$

where we set $\psi_c = 2\pi W$ for later convenience. This matrix is the convolution matrix arising from the impulse response of the ideal lowpass filter with cutoff ψ_c.

The problem of maximum energy concentration is to find that finite sequence $\mathbf{a}$ that maximizes the performance index $\mathcal{J}$. This problem has been studied extensively both for discrete and continuous time sequences, see [1095] for a nice review. The maximization of the Rayleigh quotient is realized by the maximum eigenvector of the prolate matrix A, that is, the eigenvector belonging to the maximum eigenvalue, say, λ_0:

$$A\mathbf{a} = \lambda_0 \mathbf{a} \tag{20.11.5}$$

The prolate matrix is notoriously ill-conditioned having approximately $2NW$ eigenvalues that are very near one, and the remaining eigenvalues decreasing rapidly to zero. The following table lists the eigenvalues in decreasing order for the case $N = 21$ and $W = 0.2$, so that $2NW = 8.4$, its condition number being, $\text{cond}(A) = 5.1063\times10^{16}$:

i	λ_i	i	λ_i
0	0.99999999998517786000	11	0.00131552671490021500
1	0.99999999795514627000	12	0.00007986915605618046
2	0.99999987170540139000	13	0.00000365494381482577
3	0.99999517388508363000	14	0.00000012731149204486
4	0.99987947149714795000	15	0.00000000336154097643
5	0.99792457099956200000	16	0.00000000006621668668
6	0.97588122145542644000	17	0.00000000000094327944
7	0.83446090480119717000	18	0.00000000000000920186
8	0.45591142240913063000	19	0.00000000000000004034
9	0.11887181858959120000	20	0.00000000000000001958
10	0.01567636516215985600		

These were generated by the following MATLAB code:

```
N = 21; W = 0.2;
n = 0:N-1; f = 2*W*sinc(2*W*n);
A = toeplitz(f,f);
lambda = svd(A);
```

The eigenvectors of the prolate matrix are referred to as the *discrete prolate spheroidal sequences* (DPSS), and the first $2NW$ of them are relevant in multitaper methods of spectral analysis [1100] .

For the array problem, we are interested only in the maximum eigenvector. A simple way to compute it is by the power iteration, that is, $\mathbf{a}_n = A\mathbf{a}_{n-1} = A^n\mathbf{a}_0$. However, because the corresponding eigenvalue λ_0 and the next highest one are so close to unity, the iteration will be very slow converging.

A more efficient approach is to apply the inverse power iteration on the matrix $Q = I - A$, that is, $\mathbf{a}_{n+1} = Q^{-1}\mathbf{a}_n = Q^{-n}\mathbf{a}_0$. This iteration converges to the minimum eigenvector of Q, which is the same as the maximum eigenvector of A. The minimum eigenvalue of Q is $1 - \lambda_0$, which is very small and its inverse $(1 - \lambda_0)^{-1}$ very large, causing the iteration to converge very fast.

For the array problem one needs to know the relationship of the bandwidth parameter W to the desired sidelobe level R and the array length N. Because the Taylor-Kaiser window is a good approximation to the maximum eigenvector, one expects to have a relationship among the parameters W, B, N, R. As pointed out by Kaiser and Walden [1089,1097], this relationship is approximately $W = B/N$.

We have improved this relationship slightly by using the results of [1101] to arrive at the following empirical formula, which works well over the range $14 \le R \le 120$ dB:

$$\boxed{W = \frac{0.95B + 0.14}{N}} \tag{20.11.6}$$

This leads to the following design procedure. Given N and R, we calculate B using the function `taylorbw`, described in Sec. 20.10, then calculate W from (20.11.6), and construct the prolate matrix A, and $Q = I - A$, then, apply the inverse power iteration initialized with the unit vector $\mathbf{a}_0 = [1,0,0,\dots,0]^T$:

```
$\mathbf{a} = [1,0,0,\dots,0]^T$
for $n = 1,2,\dots,N_{\text{iter}}$,
    $\mathbf{a} = Q^{-1}\mathbf{a}$
```

The algorithm is insensitive to the choice of the initial vector $\mathbf{a}_0$ and converges extremely fast, requiring about 1–3 iterations (we use 3 by default).

We determine the 3-dB width by simply equating it to that of the Taylor array, that is, $\Delta\psi = 2\pi\Delta u/N$, where Δu is also obtained from the function `taylorbw`. Even so, the prolate array's mainlobe, as a whole, is slightly narrower than that of the Taylor array. The MATLAB function `prol` implements the above procedure:

```
[a, dph] = prol(d,ph0,N,R);        % prolate array
```

The function `prolmat` constructs the prolate matrix for given N, W:

```
A = prolmat(N,W);             % prolate matrix
```

Fig. 20.11.1 shows a design example with $N = 21$ and $R = 30$. The left graph plots the array patterns $|A(\psi)|$ for the prolate and Taylor designs at broadside. The right graph shows the same prolate array steered towards 60°. The 3-dB width is also indicated on the figure. We note that the Taylor array has a slightly wider mainlobe and slightly lower sidelobes, whereas the prolate design meets the sidelobe specification exactly. The graphs were computed by the following MATLAB code:

```
d = 0.5; N = 21; R = 30; ph0 = 90;

[at,Dt] = taylor1p(d,ph0,N,R);            % Taylor design
[ap,Dp] = prol(d,ph0,N,R);                % Prolate design

f = linspace(-1,1,1001); psi = pi*f;      % normalized wavenumber ψ

At = 20*log10(abs(dtft(at,-psi))); At = At-max(At);     % compute pattern
Ap = 20*log10(abs(dtft(ap,-psi))); Ap = Ap-max(Ap);

figure; plot(f,Ap, '-', f,At,'--');

ap = steer(d,ap,60);                   % redesign steered towards 60°
[gp,phi] = gain1d(d,ap,720);           % normalized gain

figure; dbz(phi,gp);
```

Fig. 20.11.2 compares the corresponding array weights of the prolate and Taylor designs for the cases $N = 21$ and $N = 41$, and $R = 30$ dB.

20.12 Taylor Line Source

Taylor's ideal line source pattern [1086], given in Eq. (20.6.21), has a mainlobe when $|u| \le A$ and equiripple sidelobes when $|u| \ge A$. The equiripple behavior arises from the fact that the pattern switches to its cosine form when $|u| \ge A$, as shown below:

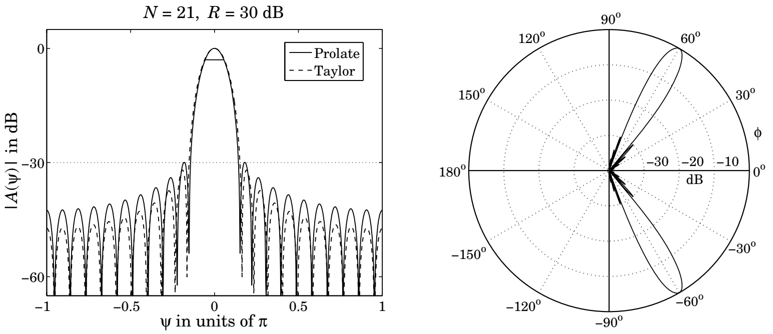


Fig. 20.11.1 Prolate array design.

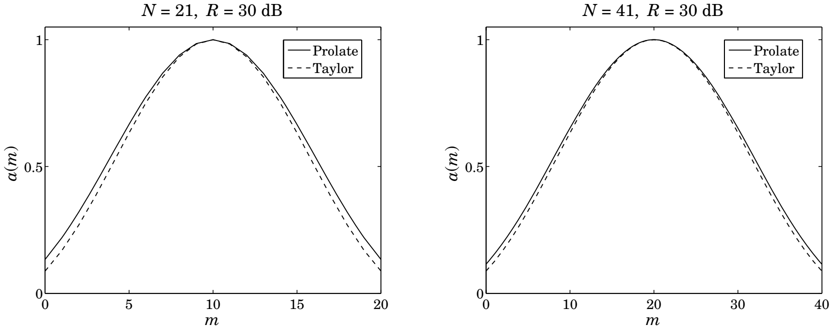
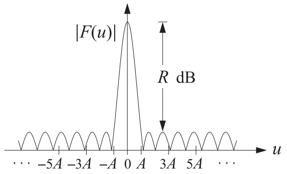


Fig. 20.11.2 Comparison of prolate and Taylor array weights.



$$F(u) = \begin{cases} \cosh\left(\pi\sqrt{A^2 - u^2}\right), & \text{if } |u| \le A \\ \cos\left(\pi\sqrt{u^2 - A^2}\right), & \text{if } |u| \ge A \end{cases} \tag{20.12.1}$$

The sidelobe level (in absolute units) is the ratio of the mainlobe peak height $|F(0)| = \cosh(\pi A)$ to the sidelobe height, which is unity:

$$R_a = \cosh(\pi A) \quad \Rightarrow \quad A = \frac{1}{\pi}\text{acosh}(R_a) \tag{20.12.2}$$

The pattern $F(u)$ may be thought of as a limiting form of the Chebyshev array when the number N of array elements becomes large [1074]. To see this, we consider the Chebyshev polynomial $T_N(x) = \cosh(N\xi)$, where $x = \cosh\xi$. In the limit of large N

and small ξ, with the product $N\xi$ kept constant, we can use the approximation

$$x=\cosh\xi\simeq 1+\frac{1}{2}\xi^2 \quad\Rightarrow\quad \xi\simeq\sqrt{2x-2}\quad\Rightarrow\quad T_N(x)\simeq\cosh\left(N\sqrt{2x-2}\right) \tag{20.12.3}$$

For an N-element array, the Chebyshev pattern is defined by Eqs. (20.9.2) and (20.9.3):

$$A(\psi)=T_{N-1}\left(x_0\cos\frac{\psi}{2}\right),\quad R_a=T_{N-1}(x_0) \tag{20.12.4}$$

Recalling that $\psi=2\pi u/N$, it follows that ψ will be small in the limit of large N and fixed u, thus, we may apply the approximation $\cos(\psi/2)\simeq 1-\psi^2/8$. Then, using the Chebyshev approximation (20.12.3), we have:

$$\begin{aligned}A(\psi)&\simeq\cosh\left((N-1)\sqrt{2x_0\cos\frac{\psi}{2}-2}\right)=\cosh\left((N-1)\sqrt{2x_0\left(1-\frac{\psi^2}{8}\right)-2}\right)\\&=\cosh\left(\sqrt{(N-1)^2(2x_0-2)-\left(\frac{(N-1)\psi}{2}\right)^2}\right)\end{aligned}$$

We also have $R_a=T_{N-1}(x_0)\simeq\cosh\left((N-1)\sqrt{2x_0-2}\right)$. Comparing with (20.12.2), we may identify $\pi A=(N-1)\sqrt{2x_0-2}$. We also note that in the large-N limit:

$$\frac{(N-1)\psi}{2}=\frac{(N-1)2\pi u}{2N}\simeq\pi u$$

It follows that the limiting form of $A(\psi)$ is precisely the pattern (20.12.1).

Taylor introduced a modification of the ideal pattern so the first few sidelobes, say the first $\bar{n}$, are essentially equiripple at the given sidelobe level R, while the remaining ones follow the $\sin\pi u/\pi u$ attenuation rate of the uniform array. The method essentially preserves the mainlobe width and sidelobe level of the Chebyshev array, while allowing the far sidelobes to decay faster.

The zeros of the $\sin\pi u/\pi u$ pattern occur at the integers $u_n=\pm n$, $n=1,2,\dots$, whereas the zeros of the ideal pattern (20.12.1) occur at the locations:

$$\sqrt{u_n^2-A^2}=(n-0.5)\quad\Rightarrow\quad u_n=\pm\sqrt{A^2+(n-0.5)^2},\quad n=1,2,\dots \tag{20.12.5}$$

Taylor defined a new pattern such that its zeros are:

$$u_n=\begin{cases}\pm\sigma\sqrt{A^2+(n-0.5)^2}, & \text{for } n=1,2,\dots,\bar{n}-1\\ \pm n, & \text{for } n\geq\bar{n}\end{cases} \tag{20.12.6}$$

The scale parameter σ is selected to allow a smooth transition between the two sets of zeros, that is, requiring the matching condition:

$$\sigma\sqrt{A^2+(\bar{n}-0.5)^2}=\bar{n}\quad\Rightarrow\quad\sigma=\frac{\bar{n}}{\sqrt{A^2+(\bar{n}-0.5)^2}} \tag{20.12.7}$$

The 3-dB width of the ideal pattern is obtained from the condition:

$$\cosh\left(\pi\sqrt{A^2-u_3^2}\right)=\frac{1}{\sqrt{2}}\cosh(\pi A)\quad\Rightarrow\quad\Delta u=2u_3 \tag{20.12.8}$$

The 3-dB width of the modified pattern is σ times larger:

$$\Delta u=2u_3\sigma=2\sigma\sqrt{A^2-\frac{1}{\pi^2}\text{acosh}^2\left(\frac{1}{\sqrt{2}}\cosh(\pi A)\right)} \tag{20.12.9}$$

As discussed by Taylor [1086], the minimum acceptable value for $\bar{n}$ should be such that $\partial\sigma/\partial\bar{n}<0$. This gives the constraint:

$$\bar{n}\geq 2A^2+1 \tag{20.12.10}$$

Eqs. (20.12.6)–(20.12.10) define completely the properties of the modified pattern. An N-element array approximating Taylor's modified pattern can be designed by the procedure outlined in Eqs. (20.6.16) and (20.6.17), that is, selecting the first $N-1$ zeros of the continuous pattern as the zeros of the discrete pattern.

In particular, given the parameters $N, R, \bar{n}$, we calculate A from (20.12.2) and σ from (20.12.7), and define the $N-1$ conjugate zeros:

$$u_n=\begin{cases}\sigma\sqrt{A^2+(n-0.5)^2}, & \text{for } 1\leq n\leq\bar{n}-1\\ n, & \text{for } \bar{n}\leq n\leq N-\bar{n}\\ -\sigma\sqrt{A^2+(N-n-0.5)^2}, & \text{for } N-\bar{n}+1\leq n\leq N-1\end{cases} \tag{20.12.11}$$

Then, we define the array's zeros in ψ-space and z-domain:

$$\psi_n=\frac{2\pi u_n}{N},\quad z_n=e^{j\psi_n},\quad n=1,2,\dots,N-1 \tag{20.12.12}$$

and convolve them to get the array pattern polynomial:

$$A(z)=\prod_{n=1}^{N-1}(z-z_n) \tag{20.12.13}$$

The first and last $\bar{n}-1$ zeros are conjugate pairs by construction; the middle ones come in conjugate pairs because for each n in the range $\bar{n}\leq n\leq N-\bar{n}$, the integer $N-n$ is also in the same range and has a conjugated zero: $z_{N-n}=e^{j2\pi(N-n)/N}=e^{-j2\pi n/N}=z_n^*$. An exception is in the case when N is even, for which $n=N/2$ corresponds to a real zero.

The 3-dB width of the array is calculated from $\Delta\psi=2\pi\Delta u/N$. This design method is implemented by the MATLAB function `taylornb` with usage:

```
[a,dph] = taylornb(d,ph0,N,R,nbar);      % Taylor's n-bar line source array design
```

Fig. 20.12.1 shows two design examples. The left graph has $N=21$, $R=30$ dB, and $\bar{n}=5$, while the right graph has $N=41$, $R=40$ dB, and $\bar{n}=10$. The required minimum values for $\bar{n}$, calculated from (20.12.10), were $\bar{n}=4.48$ and $\bar{n}=6.69$, respectively.

In order for the middle range of ns in (20.12.11) to be nontrivial, we must necessarily have $\bar{n}\leq N/2$, which combined with the restriction (20.12.10) implies a minimum value for the array length:

$$N\geq 4A^2+2 \tag{20.12.14}$$

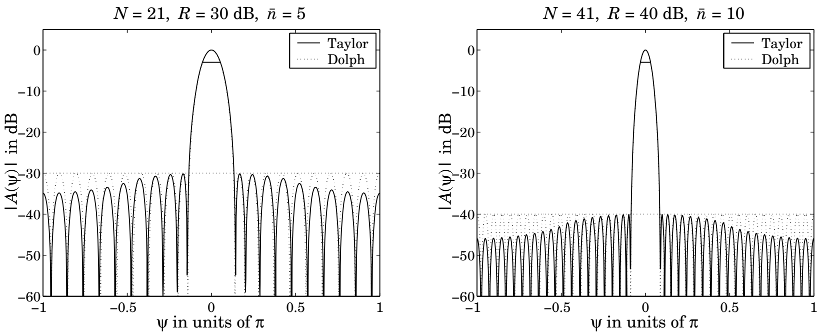


Fig. 20.12.1 Taylor $\bar{n}$ line source array design.

The following table lists some representative values of the minimum N:

R dB	$\bar{n}$	N	R dB	$\bar{n}$	N
15	2	4	50	9	18
20	2	5	55	11	22
25	3	7	60	12	25
30	4	8	65	14	29
35	5	11	70	16	33
40	6	13	75	18	37
45	7	15	80	20	41

20.13 Villeneuve Arrays

Taylor's $\bar{n}$ pattern was based on the ideal continuous line source distribution (20.12.1), which was the limit of a Dolph-Chebyshev array. The design of an N-element array was accomplished by the usual method of mapping $N-1$ continuous-case zeros to the $N-1$ zeros of the array.

Villeneuve [1102] introduced an alternative design method whose starting point was a true N-element Chebyshev array, instead of the ideal limiting form. The modified array was designed by choosing its first (and last) $\bar{n}-1$ zeros to coincide with the (stretched) zeros of the Chebyshev array, and the remaining zeros to coincide with zeros of an N-element uniform array.

We recall from Sec. 20.9 that the $N-1$ zeros of and N-element Chebyshev array are constructed by:

$$x_n = \cos\left(\frac{(2n-1)\pi}{2(N-1)}\right), \quad \psi_n = 2\,\mathrm{acos}\left(\frac{x_n}{x_0}\right), \quad z_n = e^{j\psi_n} \tag{20.13.1}$$

for $n = 1, 2, \dots, N-1$, where x_0 is determined by $T_N(x_0) = R_a$, and R_a is the sidelobe level in absolute units. Villeneuve modified the above zeros as follows:

$$\psi_n = \begin{cases} +\sigma\, 2\,\mathrm{acos}\left(\dfrac{x_n}{x_0}\right), & \text{for } 1 \le n \le \bar{n}-1 \\[2ex] \dfrac{2\pi n}{N}, & \text{for } \bar{n} \le n \le N-\bar{n} \\[2ex] -\sigma\, 2\,\mathrm{acos}\left(\dfrac{x_{N-n}}{x_0}\right), & \text{for } N-\bar{n}+1 \le n \le N-1 \end{cases} \tag{20.13.2}$$

where x_n are given as in (20.13.1). By construction, we have $\psi_{N-n} = -\psi_n$, for $1 \le n \le \bar{n}-1$, which implies that the first and last $\bar{n}$ zeros are conjugate pairs.

The scale factor σ is fixed by requiring a smooth transition between the two sets of zeros at $n = \bar{n}$, that is,

$$\sigma\, 2\,\mathrm{acos}\left(\frac{x_{\bar{n}}}{x_0}\right) = \frac{2\pi\bar{n}}{N}, \quad \text{where} \quad x_{\bar{n}} = \cos\left(\frac{(2\bar{n}-1)\pi}{2(N-1)}\right) \tag{20.13.3}$$

With $z_n = e^{j\psi_n}$, the array polynomial is then formed by

$$A(z) = \prod_{n=1}^{N-1}(z - z_n)$$

The 3-dB width of the new design is taken to be σ times greater than that of the Dolph-Chebyshev case. The MATLAB function `ville` implements this method:

```
[a,dph] = ville(d,ph0,N,R,nbar);      % Villeneuve array design
```

The method applied to the two examples of Fig. 20.12.1 produces virtually identical graphs, and we do not repeat them here.

20.14 Multibeam Arrays

An array can form multiple narrow beams towards different directions. For example, suppose it is desired to form three beams towards the steering angles ϕ_1, ϕ_2, and ϕ_3. The weights for such a multibeam array can be obtained by superimposing the weights of a single broadside array, say $w(m)$, steered towards the three angles. Defining the corresponding scanning phases $\psi_i = kd\cos\phi_i$, $i = 1, 2, 3$, we have:

$$a(m) = A_1 e^{-jm\psi_1} w(m) + A_2 e^{-jm\psi_2} w(m) + A_3 e^{-jm\psi_3} w(m)$$

where $m = 0, \pm 1, \pm 2, \dots, \pm M$ and we assumed an odd number of array elements $N = 2M+1$. The complex amplitudes A_1, A_2, A_3 represent the relative importance of the three beams. The corresponding array factor becomes:

$$A(\psi) = A_1 W(\psi - \psi_1) + A_2 W(\psi - \psi_2) + A_3 W(\psi - \psi_3)$$

and will exhibit narrow peaks towards the three steering angles. More generally, we can form L beams towards the angles ϕ_i, $i = 1, 2, \dots, L$ by superimposing the steered beams:

$$\boxed{a(m) = \sum_{i=1}^{L} A_i e^{-jm\psi_i} w(m)}, \qquad m = 0, \pm 1, \pm 2, \dots, \pm M \tag{20.14.1}$$

where $\psi_i = kd\cos\phi_i$, $i = 1, 2, \dots, L$. For an even number of array elements, $N = 2M$, we replace Eq. (20.14.1) with:

$$\boxed{a(\pm m) = \sum_{i=1}^{L} A_i e^{\mp j(m-1/2)\psi_i} w(\pm m)}\,, \qquad m = 1, 2, \dots, M \tag{20.14.2}$$

For either even or odd N, the corresponding array factor will be the superposition:

$$\boxed{A(\psi) = \sum_{i=1}^{L} A_i W(\psi - \psi_i)} \qquad \text{(multi-beam array factor)} \tag{20.14.3}$$

The basic broadside array weights $w(m)$ can be designed to achieve a desired sidelobe level or beam width. As the broadside beam $w(m)$ is steered away from 90°, the beamwidths will broaden. To avoid grating lobes, the element spacing d must be less the quantity d_0 (and greater than $d_0/2$):

$$d_0 = \min_i d_i\,, \quad \text{where } d_i = \frac{\lambda}{1 + |\cos\phi_i|}\,, \quad i = 1, 2, \dots, L$$

This minimum is realized at the beam angle closest to endfire. If the steering angles are closer to each other than about one 3-dB beamwidth, the mainlobes will begin to merge with each other reducing the resolvability of the individual beams. This behavior is analogous to the problem of frequency resolution of multiple sinusoids.

The MATLAB function `multbeam.m` of Appendix I implements Eqs. (20.14.1) and (20.14.2). Its inputs are the vector of broadside array weights **w**—which can be designed beforehand using for example `dolph2` or `taylor1p`—and the beam angles and amplitudes ϕ_i, A_i.

Example 20.14.1: Fig. 20.14.1 shows the gains of two 21-element three-beam arrays with half-wavelength spacing, and steered towards the three angles of 45°, 90°, and 120°. The broadside array was designed as a Taylor-Kaiser array with sidelobe level of $R = 20$ and $R = 30$ dB.

The relative amplitudes of the three beams were equal to unity. The MATLAB code used to generate the right figure was:

```
w = taylor1p(0.5, 90, 21, 30);                     % unsteered weights
a = multbeam(0.5, w, [1,1,1], [45, 90, 120]);     % equal-amplitude beams
[g, ph] = array(d, a, 400);                        % compute gain
dbz(ph, g);                                        % plot gain in dB
addray(45); addray(-45);                           % add ± 45° grid rays
```

We note the broadening of the beam widths of the larger beam angles. The left array has narrower mainlobes than the right one because its sidelobe attenuation is less. But, it also exhibits more constructive interference between mainlobes causing somewhat smaller sidelobe attenuations than the desired one of 20 dB. □

Equations (20.14.1) and (20.14.2) generalize the Woodward-Lawson frequency sampling design equations (20.5.6) and (20.5.7) in the sense that the steering phases ψ_i can be arbitrary and do not have to be the DFT frequencies.

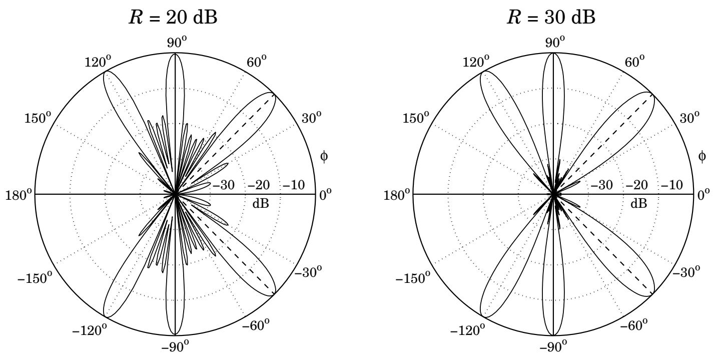


Fig. 20.14.1 Multi-beam arrays with $R = 20$ and $R = 30$ dB sidelobes.

However, if the ψ_i are chosen to be the DFT frequencies given by Eq. (20.5.1) or (20.5.5), and the broadside array is chosen to be a length-N uniform array, $w(m) = 1$, then the inverse DFT expressions (20.5.6) and (20.5.7) can be thought of as defining N beams—called the Woodward-Lawson-Butler beams—steered towards the DFT angles $\phi_i = \text{acos}(\psi_i/kd)$, that is, towards

$$\phi_i = \text{acos}\left(\frac{\psi_i}{kd}\right) = \text{acos}\left(\frac{2\pi i}{Nkd}\right) = \text{acos}\left(\frac{\lambda i}{Nd}\right)\,, \quad i = 0, 1, \dots, N-1 \tag{20.14.4}$$

The array weights will be given then by the inverse DFT:

$$a(m) = \frac{1}{N}\sum_{i=0}^{N-1} A(\psi_i)\,e^{-jm\psi_i} \tag{20.14.5}$$

and the corresponding array factor by:

$$A(\psi) = \frac{1}{N}\sum_{i=0}^{N-1} A(\psi_i)\,W(\psi - \psi_i) \tag{20.14.6}$$

where $W(\psi) = \sin(N\psi/2)/\sin(\psi/2)$ is the array factor of the uniform window. The DFT values are identified as the relative beam weights $A_i = A(\psi_i)/N$.

A single Butler beam, say the jth beam, can be turned on by choosing $A_i = \delta_{ij}$. By successively turning on the Butler beams one by one, the array will act as a scanning array. Fig. 20.14.2 depicts such a multi-beam array structure. The inverse DFT box implements Eq. (20.14.5). The inputs are the "beams" A_i and the outputs are the weights $a(m)$.

Somewhat before the advent of the FFT algorithm, Butler proposed a hardware realization of the inverse DFT network, which was quickly recognized to be equivalent to the FFT algorithm [1103,1104,1106–1109]. The DFT matrix realization of this network is called the Blass matrix in the antenna array context [18,10].

Example 20.14.2: Fig. 20.14.3 shows the individual Butler beams turned on successively for an eight-element array. Both the standard and alternative DFT frequency sets are shown.

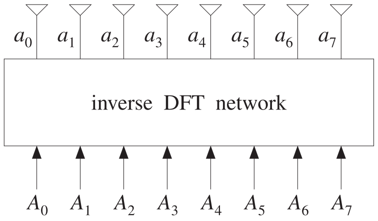
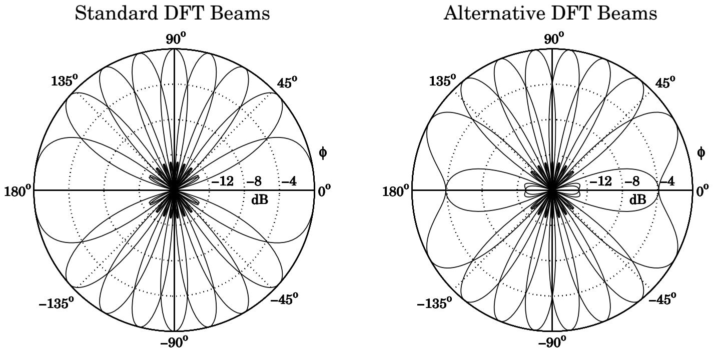

a_0 a_1 a_2 a_3 a_4 a_5 a_6 a_7
inverse DFT network
A_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7

Fig. 20.14.2 Woodward-Lawson-Butler beam matrix network for $N = 8$.

There are eight beams in each graph. For the standard DFT set, the two endfire beams count as one, that is, the $i = 0$ beam.

The sidelobes are at the 13-dB level because these are scanned versions of the uniform array. The mainlobes intersect exactly half-way between the DFT frequencies ψ_i, that is, the ith beam intersects the neighboring ones at $\psi = \psi_i + \pi/N = 2\pi(i+0.5)/N$. These intersection points are approximately 4 dB down (3.92 dB to be exact) from the main peaks. The 4-dB gain circle intersects the gain curves at these points. □


Standard DFT Beams

Alternative DFT Beams

Fig. 20.14.3 Woodward-Lawson-Butler beams for $N = 8$.

20.15 Problems

20.1 *Computer Experiment—Taylor's one-parameter/$\bar{n}$ array design.* Taylor's $\bar{n}$ distribution of Sec. 20.12 can also be applied to Taylor's one-parameter continuous distribution of Sec. 20.10. First, show that the zeros of Eq. (20.10.2) occur at

$$u_n = \sqrt{B^2 + n^2}\,, \quad n = 1, 2, \dots$$

Then, choose $\bar{n}$ such that $\bar{n} \le N/2$ and define an N-element array by its $N-1$ zeros:

$$u_n = \begin{cases} \sigma\sqrt{B^2 + n^2}\,, & \text{for } 1 \le n \le \bar{n} - 1 \\ n\,, & \text{for } \bar{n} \le n \le N - \bar{n} \\ -\sigma\sqrt{B^2 + (N-n)^2}\,, & \text{for } N - \bar{n} + 1 \le n \le N - 1 \end{cases} \tag{20.15.1}$$

Fix the parameter σ such that $\sigma\sqrt{B^2 + \bar{n}^2} = \bar{n}$. Then, define the array polynomial:

$$A(z) = \prod_{n=1}^{N-1}(z - z_n)\,, \quad \psi_n = \frac{2\pi u_n}{N}\,, \quad z_n = e^{j\psi_n}\,, \quad n = 1, 2, \dots, N-1 \tag{20.15.2}$$

Write a MATLAB function that implements this procedure, and takes as input the parameters $N, R, \bar{n}$ and outputs the array weights and 3-dB width.

Apply your function to the following example $N = 21$, $R = 30$ dB, $\bar{n} = 5$, with half-wavelength spacing $d = \lambda/2$. You will notice that, like the prolate array, the mainlobe is slightly narrower and the sidelobe level slightly better matched than the Taylor-Kaiser array. On the same graph, plot the array patterns $|A(\psi)|$ in dB for the present design, the Taylor Kaiser and the prolate arrays designed with the same specifications. Vary $\bar{n}$ to understand its effect on the design.

20.2 *Computer Experiment—Villeneuve array design.* Redesign the examples shown in Fig. 20.12.1 using Villeneuve's array design method and plot the array responses together with those of that figure. Vary the parameters $N, R, \bar{n}$ and compare the range of similarity of the Villeneuve versus the Taylor $\bar{n}$ method.

21

Currents on Linear Antennas

21.1 Hallén and Pocklington Integral Equations

In Sec. 14.4, we determined the electromagnetic fields generated by a given current distribution on a thin linear antenna, but did not discuss the mechanism by which the current distribution is set up and maintained. In Chap. 16, we assumed that the currents were sinusoidal, but this was only an approximation. Here, we discuss the integral equations that determine the exact form of the currents.

An antenna, whether transmitting or receiving, is always driven by an external source field. In transmitting mode, the antenna is driven by a generator voltage applied to its input terminals, and in receiving mode, by an incident electric field (typically, a uniform plane wave if it is arriving from far distances.) In either case, we will refer to this external source field as the "incident" field $\boldsymbol{E}_{\text{in}}$.

The incident field $\boldsymbol{E}_{\text{in}}$ induces a current on the antenna. In turn, the current generates its own field $\boldsymbol{E}$, which is radiated away. The total electric field is the sum $\boldsymbol{E}_{\text{tot}} = \boldsymbol{E} + \boldsymbol{E}_{\text{in}}$. Assuming a perfectly conducting antenna, the *boundary conditions* are that the tangential components of the total electric field vanish on the antenna surface. These boundary conditions are enough to determine the current distribution induced on the antenna.

Fig. 21.1.1 depicts a z-directed thin cylindrical antenna of length l and radius a, with a current distribution $I(z)$ along its length. We will concentrate only on the z-component E_z of the electric field generated by the current and use cylindrical coordinates.

For a perfectly conducting antenna, the current is essentially a surface current at radial distance $\rho = a$ with surface density $\boldsymbol{J}_s(z) = \hat{\mathbf{z}}\, I(z)/2\pi a$, where in the "thin-wire approximation," we may assume that the density is azimuthally symmetric with no dependence on the azimuthal angle ϕ. The corresponding volume current density will be as in Eq. (14.4.2):

$$\boldsymbol{J}(\mathbf{r}) = \boldsymbol{J}_s(z)\,\delta(\rho - a) = \hat{\mathbf{z}}\, I(z)\,\delta(\rho - a)\,\frac{1}{2\pi a} \equiv \hat{\mathbf{z}}\, J_z(\mathbf{r})$$

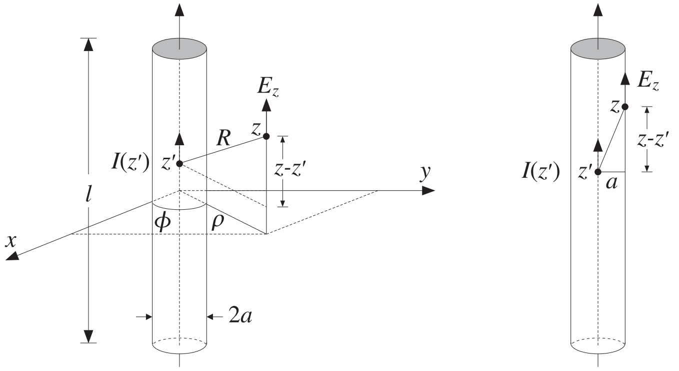


Fig. 21.1.1 Thin-wire model of cylindrical antenna.

Following the procedure of Sec. 14.4, we obtain the z-component of the vector potential:

$$\begin{aligned}
A_z(z,\rho,\phi) &= \frac{\mu}{4\pi}\int_{V'}\frac{J_z(\mathbf{r}')e^{-jkR}}{R}\,d^3\mathbf{r}' = \frac{\mu}{4\pi}\int_{V'}\frac{I(z')\,\delta(\rho'-a)\,e^{-jkR}}{2\pi aR}\,\rho'd\rho'd\phi'dz' \\
&= \frac{\mu}{4\pi}\int_{-l/2}^{l/2}\int_0^{2\pi}\frac{I(z')\,e^{-jkR}}{2\pi R}\,d\phi'dz'
\end{aligned}$$

where $R = |\mathbf{r}-\mathbf{r}'| = \sqrt{(z-z')^2+|\boldsymbol{\rho}-\boldsymbol{\rho}'|^2}$. Because $\rho' = a$, we have:

$$|\boldsymbol{\rho}-\boldsymbol{\rho}'|^2 = \rho^2+a^2-2\boldsymbol{\rho}\cdot\boldsymbol{\rho}' = \rho^2+a^2-2\rho a\cos(\phi'-\phi)$$

and because ϕ' appears only through the difference $\phi'-\phi$, we may change the variable of integration from ϕ' to $\phi'-\phi$. This implies that A_z will be cylindrically symmetric, that is, independent of ϕ. It follows that:

$$A_z(z,\rho)=\frac{\mu}{4\pi}\int_{-l/2}^{l/2}I(z')\,G(z-z',\rho)\,dz' \tag{21.1.1}$$

where we defined the *exact thin-wire kernel*:

$$\boxed{G(z-z',\rho)=\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-jkR}}{R}\,d\phi'} \tag{21.1.2}$$

with $R = \sqrt{(z-z')^2+\rho^2+a^2-2\rho a\cos\phi'}$. In the limit of a thin antenna, $a\to 0$, Eq. (21.1.1) reduces to:

$$A_z(z,\rho)=\frac{\mu}{4\pi}\int_{-l/2}^{l/2}I(z')\,G_{\text{app}}(z-z',\rho)\,dz' \tag{21.1.3}$$

where $G_{\text{app}}(z-z',\rho)$ is the *approximate* or *reduced thin-wire kernel*:

$$\boxed{G_{\text{app}}(z-z',\rho)=\frac{e^{-jkR}}{R}} \tag{21.1.4}$$

with $R=\sqrt{(z-z')^2+\rho^2}$. Eq. (21.1.3) is the same as (14.4.3) because the limit $a=0$ is equivalent to assuming that the current density is a line current $\boldsymbol{J}(\mathbf{r})=\hat{\mathbf{z}}\,I(z)\,\delta(x)\,\delta(y)$, as given by Eq. (14.4.1).

Given the vector potential $A_z(z,\rho)$, the z-component of the electric field generated by the current is obtained from Eq. (14.4.6):

$$j\omega\mu\epsilon\,E_z(z,\rho)=(\partial_z^2+k^2)\,A_z(z,\rho) \tag{21.1.5}$$

The values of the vector potential A_z and the electric field E_z on the surface of the wire antenna are obtained by setting $\rho=a$:

$$A_z(z,a)=\frac{\mu}{4\pi}\int_{-l/2}^{l/2}I(z')\,G(z-z',a)\,dz' \tag{21.1.6}$$

To simplify the notation, we will denote $A_z(z,a)$ and $G(z-z',a)$ by $A_z(z)$ and $G(z-z')$. The boundary condition on the surface is that the z-component of the total electric field vanish, that is, at $\rho=a$:

$$E_{z,\text{tot}}(z,a)=E_z(z,a)+E_{z,\text{in}}(z,a)=0$$

Thus, with $E_z(z)=E_z(z,a)$ and $E_{\text{in}}(z)=E_{z,\text{in}}(z,a)$, we have $E_z(z)=-E_{\text{in}}(z)$, and Eq. (21.1.5) can be expressed in terms of the z-component of the incident field:

$$\boxed{(\partial_z^2+k^2)\,A_z(z)=-j\omega\mu\epsilon\,E_{\text{in}}(z)} \tag{21.1.7}$$

Either kernel can be used in Eq. (21.1.6). If the approximate kernel $G_{\text{app}}(z)$ is used, then it is still meaningful to consider the boundary conditions at the cylindrical surface (i.e., at $\rho=a$) of the antenna, as shown on the right of Fig. 21.1.1.

To summarize, given an incident field $E_{\text{in}}(z)$ that is known along the length of the antenna, Eq. (21.1.7) may be solved for $A_z(z)$ and then the integral equation (21.1.6) can be solved for the current $I(z)$.

Depending on how this procedure is carried out, one obtains either the Hallén or the Pocklington equations. Solving Eq. (21.1.7) by formally inverting the differential operator $(\partial_z^2+k^2)$ and combining with (21.1.6), we obtain *Hallén's integral equation*:

$$\boxed{\frac{\mu}{4\pi}\int_{-l/2}^{l/2}I(z')\,G(z-z')\,dz' = -j\omega\mu\epsilon\,(\partial_z^2+k^2)^{-1}E_{\text{in}}(z)}\quad\text{(Hallén)} \tag{21.1.8}$$

Alternatively, applying the differential operator $(\partial_z^2+k^2)$ directly to Eq. (21.1.6) and combining with (21.1.7) , we obtain *Pocklington's integral equation*:

$$\boxed{\frac{\mu}{4\pi}\int_{-l/2}^{l/2}I(z')\,(\partial_z^2+k^2)\,G(z-z')\,dz' = -j\omega\mu\epsilon\,E_{\text{in}}(z)}\quad\text{(Pocklington)} \tag{21.1.9}$$

The two integral equations must be solved subject to the constraint that the current $I(z)$ vanish at the antenna ends, that is, $I(l/2)=I(-l/2)=0$. The exact and approximate kernels evaluated on the antenna surface are:

$$\begin{aligned}
G(z-z') &= \frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-jkR}}{R}\,d\phi', \quad R=\sqrt{(z-z')^2+2a^2-2a^2\cos\phi'} \\
G_{\text{app}}(z-z') &= \frac{e^{-jkR}}{R}, \quad R=\sqrt{(z-z')^2+a^2}
\end{aligned} \tag{21.1.10}$$

The inverse differential operator in the right-hand side of Eq. (21.1.8) can be rewritten as an integral convolutional operator acting on E_{in}. We discuss this in detail in Sec. 21.3. We will then consider the numerical solutions of these equations using either the exact or the approximate kernels. The numerical evaluation of these kernels is discussed in Sec. 21.7.

21.2 Delta-Gap, Frill Generator, and Plane-Wave Sources

Although the external source field $E_{\text{in}}(z)$ can be specified arbitrarily, there are two special cases of practical importance. One is the so-called delta-gap model, which imitates the way a transmitting antenna is fed by a transmission line. The other is a uniform plane wave incident at an angle on a receiving antenna connected to a load impedance. Fig. 21.2.1 depicts these cases.

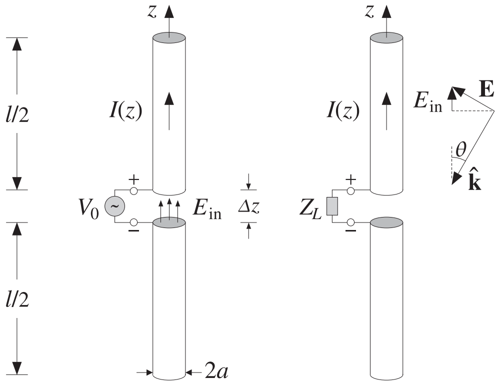


Fig. 21.2.1 External sources acting on a linear antenna.

The left figure shows the *delta-gap* model of a generator voltage applied between the upper and lower halves of the antenna across a short gap of length Δz. The applied voltage V_0 can be thought of as arising from an electric field—the "incident" field in this case—which exists only within the gap, such that

$$V_0 = \int_{-\Delta z/2}^{\Delta z/2} E_{\text{in}}(z)\,dz \tag{21.2.1}$$

A simplified case arises when we take the limit $\Delta z \to 0$. Then, approximately, $V_0 = E_{\text{in}}\Delta z$, or $E_{\text{in}} = V_0/\Delta z$. In order to maintain a finite value of V_0 in the left-hand side of Eq. (21.2.1), E_{in} must become commensurately large. This means that in this limit,

$$\boxed{E_{\text{in}}(z) = V_0\delta(z)} \qquad \text{(delta-gap model of incident field)} \tag{21.2.2}$$

King [3] has discussed the case of a finite Δz. An alternative type of excitation input is the *frill generator* [6,7] defined by:

$$E_{\text{in}}(z) = \frac{V_0}{2\ln(b/a)}\left[\frac{e^{-jkR_a}}{R_a} - \frac{e^{-jkR_b}}{R_b}\right], \qquad \begin{aligned} R_a &= \sqrt{z^2 + a^2} \\ R_b &= \sqrt{z^2 + b^2} \end{aligned} \tag{21.2.3}$$

where $b > a$. The case of a receiving antenna with a uniform plane wave incident at a polar angle θ and such that the propagation vector $\hat{\mathbf{k}}$ is co-planar with the antenna axis is shown on the right of Fig. 21.2.1.

The electric field vector is perpendicular to $\hat{\mathbf{k}}$ and has a space dependence $E_0 e^{-j\mathbf{k}\cdot\mathbf{r}}$. For a thin antenna, we may evaluate the field along the z-axis, that is, we set $x = y = 0$ so that $e^{-j\mathbf{k}\cdot\mathbf{r}} = e^{-jk_z z} = e^{jkz\cos\theta}$ because $k_z = -k\cos\theta$. Then, the z-component of the incident field will be:

$$\boxed{E_{\text{in}}(z) = E_0\sin\theta\, e^{jkz\cos\theta}} \qquad \text{(incident uniform plane wave)} \tag{21.2.4}$$

If the wave is incident from broadside ($\theta = \pi/2$), then $E_{\text{in}}(z) = E_0$, that is, a constant along the antenna length. And, if $\theta = 0$ or π, then $E_{\text{in}}(z) = 0$.

21.3 Solving Hallén's Equation

Instead of working with the vector potential $A_z(z)$ it proves convenient to work with a scaled version of it that has units of volts and is defined as:

$$V(z) = 2jcA_z(z) \tag{21.3.1}$$

where c is the speed of light. We note that $V(z)$ is not the scalar potential $\varphi(z)$ along the antenna length. From the Lorenz condition, Eq. (14.4.5), we have $\partial_z A_z = -j\omega\mu\epsilon\varphi(z)$. Multiplying by $2jc$ and noting that $c\omega\epsilon\mu = \omega/c = k$, we find:

$$\partial_z V(z) = 2k\varphi(z) \tag{21.3.2}$$

Multiplying both sides of Eq. (21.1.7) by $2jc$, we can rewrite it as:

$$\boxed{(\partial_z^2 + k^2)V(z) = 2kE_{\text{in}}(z)} \tag{21.3.3}$$

Similarly, Eq. (21.1.6) becomes:

$$\boxed{\frac{j\eta}{2\pi}\int_{-h}^{h} G(z - z')I(z')\,dz' = V(z)} \tag{21.3.4}$$

where $\eta = \sqrt{\mu/\epsilon}$, and for later convenience, we introduced the half-length $h = l/2$ of the antenna. Eqs. (21.3.3)–(21.3.4) represent our rescaled version of Hallén's equations.

Formally, we can write $V(z) = 2k(\partial_z^2 + k^2)^{-1}E_{\text{in}}(z)$, but we prefer to express $V(z)$ as an integral operator acting on $E_{\text{in}}(z)$. A particular solution of (21.3.3) is obtained with the help of the Green's function $F(z)$ for this differential equation:

$$\boxed{(\partial_z^2 + k^2)F(z) = 2k\delta(z)} \tag{21.3.5}$$

The general solution of Eq. (21.3.3) is obtained by adding the most general solution of the homogeneous equation, $(\partial_z^2 + k^2)V(z) = 0$, to the Green's function solution:

$$V(z) = C_1 e^{jkz} + C_2 e^{-jkz} + \int_{-h}^{h} F(z - z')E_{\text{in}}(z')\,dz' \tag{21.3.6}$$

With a re-definition of the constants C_1, C_2, we can also write:

$$V(z) = C_1 \cos kz + C_2 \sin kz + \int_{-h}^{h} F(z-z')E_{\rm in}(z')\,dz' \tag{21.3.7}$$

In fact, $F(z)$ itself is defined up to an arbitrary solution of the homogeneous equation. If $F(z)$ satisfies Eq. (21.3.5), so does $F_1(z) = F(z) + C_1 e^{jkz} + C_2 e^{-jkz}$, with arbitrary constants C_1, C_2. Some possible choices for $F(z)$ are as follows. They differ from each other by a homogeneous term:

$$\boxed{\begin{aligned} F_1(z) &= je^{-jk|z|} = F_2(z) + j\cos kz \\ F_2(z) &= \sin k|z| = F_3(z) - \sin kz \\ F_3(z) &= 2\sin(kz)u(z) = F_4(z) + 2\sin kz \\ F_4(z) &= -2\sin(kz)u(-z) \end{aligned}} \tag{21.3.8}$$

where $u(z)$ is the unit-step function. All satisfy Eq. (21.3.5) as well as the required discontinuity conditions on their first derivative, that is,

$$F'(0+) - F'(0-) = 2k \tag{21.3.9}$$

This discontinuity condition is obtained by integrating Eq. (21.3.5) over the small interval $-\epsilon \le z \le \epsilon$ and then taking the limit $\epsilon \to 0$ and assuming that $F(z)$ itself is continuous at $z = 0$. Depending on the choice of $F(z)$, the corresponding solution $V(z)$ of Eq. (21.3.3) can be written in the equivalent forms (each with different C_1, C_2):

$$\begin{aligned} V(z) &= C_1 e^{jkz} + C_2 e^{-jkz} + \int_{-h}^{h} je^{-jk|z-z'|}E_{\rm in}(z')\,dz' \\ V(z) &= C_1 e^{jkz} + C_2 e^{-jkz} + \int_{-h}^{h} \sin\big(k|z-z'|\big)E_{\rm in}(z')\,dz' \\ V(z) &= C_1 e^{jkz} + C_2 e^{-jkz} + 2\int_{-h}^{z} \sin\big(k(z-z')\big)E_{\rm in}(z')\,dz' \\ V(z) &= C_1 e^{jkz} + C_2 e^{-jkz} - 2\int_{z}^{h} \sin\big(k(z-z')\big)E_{\rm in}(z')\,dz' \end{aligned} \tag{21.3.10}$$

We will use mostly the first and second choices for $F(z)$, that is, $F(z) = je^{-jk|z|}$ and $F(z) = \sin k|z|$. Combining the solution for $V(z)$ with Eq. (21.3.4), we obtain the equivalent form of *Hallén's integral equation for an arbitrary incident field*:

$$\boxed{\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = C_1 e^{jkz} + C_2 e^{-jkz} + \int_{-h}^{h} F(z-z')E_{\rm in}(z')\,dz'} \tag{21.3.11}$$

or, alternatively,

$$\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = C_1 \cos kz + C_2 \sin kz + \int_{-h}^{h} F(z-z')E_{\rm in}(z')\,dz'$$

The constants C_1, C_2 are determined from the end conditions $I(h) = I(-h) = 0$. Next, we consider the particular forms of Eq. (21.3.11) in the delta-gap and plane-wave cases. In the delta-gap case, we have $E_{\rm in}(z) = V_0\delta(z)$ and the integral on the right-hand side can be done trivially, giving:

$$\int_{-h}^{h} F(z-z')E_{\rm in}(z')\,dz' = \int_{-h}^{h} F(z-z')V_0\delta(z')\,dz' = V_0 F(z)$$

Thus, we have the integral equation:

$$\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = C_1 \cos kz + C_2 \sin kz + V_0 F(z)$$

We expect the current $I(z)$ to be an even function of z (because $E_{\rm in}(z)$ is), and thus we may drop the C_2 term. Using $F(z) = \sin k|z|$ as our Green's function choice, we obtain *Hallén's equation for the delta-gap case*:

$$\boxed{\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = V(z) = C_1 \cos kz + V_0 \sin k|z|} \tag{21.3.12}$$

This equation forms the basis for determining the current on a center-driven linear antenna. We will consider several approximate solutions of it as well as numerical solutions based on moment methods.

We can verify that $V(z)$ correctly gives the potential difference between the upper and lower halves of the antenna. Differentiating $V(z)$ about $z = 0$ and using Eq. (21.3.2), we have:

$$V'(0+) - V'(0-) = 2kV_0 = 2k\big(\varphi(0+) - \varphi(0-)\big) \quad \Rightarrow \quad \varphi(0+) - \varphi(0-) = V_0$$

As a second example, consider the case of an antenna receiving a uniform plane wave with incident field as in Eq. (21.2.4). Using $F(z) = je^{-jk|z|}$ as the Green's function, the convolution integral of $F(z)$ and $E_{\rm in}(z)$ can be done easily giving:

$$\int_{-h}^{h} je^{-jk|z-z'|}E_0 \sin\theta\, e^{jkz'\cos\theta}\,dz' = \frac{2E_0}{k\sin\theta}e^{jkz\cos\theta} + (\text{homogeneous terms})$$

where the last terms are solutions of the homogeneous equation, and thus, can be absorbed into the other homogeneous terms of $V(z)$. Because the current is not expected to be symmetric in z, we must keep both homogeneous terms, resulting in *Hallén's equation for a receiving antenna*:

$$\boxed{\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = V(z) = C_1 e^{jkz} + C_2 e^{-jkz} + \frac{2E_0}{k\sin\theta}e^{jkz\cos\theta}} \tag{21.3.13}$$

21.4 Sinusoidal Current Approximation

Here, we look at simplified solutions of Eq. (21.3.12), which justify the common sinusoidal assumption for the current. We work with the approximate kernel.

Inspecting the quantity $G_{\text{app}}(z-z') = e^{-jkR}/R$ in the integral equation (21.3.12), we note that as the integration variable z' sweeps past z, the denominator becomes very large, because $R = a$ at $z' = z$. Therefore, the integral is dominated by the value of the integrand near $z' = z$. We can write approximately,

$$\frac{j\eta}{2\pi}\int_{-h}^{h} G_{\text{app}}(z-z')I(z')dz' \simeq \bar{Z}(z)I(z) \simeq \bar{Z}I(z) \tag{21.4.1}$$

where $\bar{Z}(z)$ is a sort of an average value of $j\eta G_{\text{app}}(z-z')/2\pi$ in the neighborhood of $z' = z$. This quantity varies slowly with z and we may approximate it with a constant, say $\bar{Z}$. Then, Hallén's equation (21.3.12) becomes approximately:

$$\bar{Z}I(z) = V(z) = C_1\cos kz + V_0\sin k|z|$$

This shows that $I(z)$ is approximately sinusoidal. The constant C_1 is fixed by the end-condition $I(h) = 0$, which gives:

$$C_1\cos kh + V_0\sin kh = 0 \quad\Rightarrow\quad C_1 = -V_0\frac{\sin kh}{\cos kh}$$

so that $I(z)$ becomes:

$$\bar{Z}I(z) = -V_0\frac{1}{\cos kh}\big[\sin kh\cos kz - \cos kh\sin k|z|\big] = -V_0\frac{1}{\cos kh}\sin\big(k(h-|z|)\big)$$

Solving for $I(z)$, we obtain the common standing-wave expression for the current:

$$I(z) = I(0)\frac{\sin\big(k(h-|z|)\big)}{\sin kh}\,, \qquad I(0) = -\frac{V_0\sin kh}{\bar{Z}\cos kh} \tag{21.4.2}$$

where $I(0)$ is the input current at $z = 0$. The crude approximation of Eq. (21.4.1) can be refined further using King's three-term approximation discussed in Sec. 21.6. From Eq. (21.4.2), the antenna input impedance is seen to be:

$$Z_A = \frac{V_0}{I(0)} = -\bar{Z}\cot kh \tag{21.4.3}$$

21.5 Reflecting and Center-Loaded Receiving Antennas

A similar approximation to Hallén's equation can be carried out in the plane-wave case shown in Fig. 21.2.1. We distinguish three cases: (a) $Z_L = 0$, corresponding to a reflecting parasitic antenna with short-circuited output terminals, (b) $Z_L = \infty$, corresponding to open-circuited terminals, and (c) arbitrary Z_L, corresponding to a center-loaded receiving antenna. See Ref. [12] for more details on this approach.

By finding the short-circuit current from case (a) and the open-circuit voltage from case (b), we will determine the output impedance of the receiving antenna, that is, the Thevénin impedance Z_A of the model of Sec. 15.4, and show that it is equal to the input impedance (21.4.3) of the transmitting antenna, in accordance with the reciprocity principle. We will also show from case (c) that the angular gain pattern of the receiving antenna agrees with that of the transmitting one.

Starting with the short-circuited case, the approximation of Eq. (21.4.1) applied to (21.3.13) gives:

$$\bar{Z}I(z) = V(z) = C_1e^{jkz} + C_2e^{-jkz} + \frac{2E_0}{k\sin\theta}e^{jkz\cos\theta}$$

The end-point conditions $I(h) = I(-h) = 0$ provide two equations in the two unknowns C_1, C_2, that is,

$$C_1e^{jkh} + C_2e^{-jkh} + \frac{2E_0}{k\sin\theta}e^{jkh\cos\theta} = 0$$

$$C_1e^{-jkh} + C_2e^{jkh} + \frac{2E_0}{k\sin\theta}e^{-jkh\cos\theta} = 0$$

with solution:

$$C_1 = -\frac{E_0\sin\big(kh(1+\cos\theta)\big)}{k\sin\theta\sin kh\cos kh}\,, \quad C_2 = -\frac{E_0\sin\big(kh(1-\cos\theta)\big)}{k\sin\theta\sin kh\cos kh}$$

Then, the current $I(z)$ becomes:

$$I(z) = \frac{1}{\bar{Z}}\big[C_1e^{jkz} + C_2e^{-jkz} + \frac{2E_0}{k\sin\theta}e^{jkz\cos\theta}\big] \tag{21.5.1}$$

For normal incidence, $\theta = 90^{\text{o}}$, we have $C_1 = C_2$ and Eq. (21.5.1) becomes:

$$I(z) = \frac{2E_0}{\bar{Z}k\cos kh}(\cos kh - \cos kz) \tag{21.5.2}$$

For $\theta = 0$ and $\theta = \pi$, the z-component of the incident field is zero, $E_{\text{in}}(z) = 0$, and we expect $I(z) = 0$. This can be verified by carefully taking the limit of Eq. (21.5.1) at $\theta = 0, \pi$, with the seemingly diverging term $2E_0/k\sin\theta$ getting canceled.

The short-circuit current at the output terminals is obtained by setting $z = 0$ in Eq. (21.5.1):

$$I_{\text{sc}} = I(0) = \frac{1}{\bar{Z}}\big[C_1 + C_2 + \frac{2E_0}{k\sin\theta}\big]$$

Inserting the expressions for C_1, C_2, we find:

$$I_{\text{sc}} = \frac{2E_0}{\bar{Z}k\cos kh}\,\frac{\cos kh - \cos(kh\cos\theta)}{\sin\theta} \tag{21.5.3}$$

For the open-circuit case, the incident field will induce an open-circuit voltage across the gap, and therefore, the scalar potential $\varphi(z)$ will be discontinuous at $z = 0$. In addition, the current must vanish at $z = 0$. Therefore, we must apply Eq. (21.3.13) separately to the upper and lower halves of the antenna. Using $\cos kz$ and $\sin kz$ as the homogeneous terms, instead of $e^{\pm jkz}$, we have the approximation:

$$\bar{Z}I(z) = V(z) = \begin{cases} C_1\cos kz + C_2\sin kz + \dfrac{2E_0}{k\sin\theta}e^{jkz\cos\theta}, & z \geq 0 \\[2ex] D_1\cos kz + D_2\sin kz + \dfrac{2E_0}{k\sin\theta}e^{jkz\cos\theta}, & z \leq 0 \end{cases}$$

The conditions $I(0+)=I(h)=0$ and $I(0-)=I(-h)=0$ provide four equations in the four unknowns C_1, C_2, D_1, D_2. They are:

$$C_1+\frac{2E_0}{k\sin\theta}=0\,,\qquad C_1\cos kh+C_2\sin kh+\frac{2E_0}{k\sin\theta}e^{jkh\cos\theta}=0$$

$$D_1+\frac{2E_0}{k\sin\theta}=0\,,\qquad D_1\cos kh-D_2\sin kh+\frac{2E_0}{k\sin\theta}e^{-jkh\cos\theta}=0$$

with solution:

$$C_1=D_1=-\frac{2E_0}{k\sin\theta}$$

$$C_2=\frac{2E_0(\cos kh-e^{jkh\cos\theta})}{k\sin\theta\sin kh}\,,\quad D_2=-\frac{2E_0(\cos kh-e^{-jkh\cos\theta})}{k\sin\theta\sin kh}$$

The open-circuit voltage is $V_{\rm oc}=\varphi(0+)-\varphi(0-)$. Using Eq. (21.3.2), we have:

$$V'(0+)-V'(0-)=2kV_{\rm oc}=k(C_2-D_2)\qquad\Rightarrow\qquad V_{\rm oc}=\frac{1}{2}(C_2-D_2)$$

and using the solution for C_2, D_2, we find:

$$V_{\rm oc}=\frac{2E_0}{k\sin kh}\,\frac{\cos kh-\cos(kh\cos\theta)}{\sin\theta}\tag{21.5.4}$$

Having found the short-circuit current and open-circuit voltage, we obtain the corresponding output Thevénin impedance by dividing Eq. (21.5.4) and (21.5.3):

$$Z_A=-\frac{V_{\rm oc}}{I_{\rm sc}}=-\bar{Z}\cot kh\tag{21.5.5}$$

where the minus sign is due to the fact that $I_{\rm sc}$ is flowing into (instead of out of) the top antenna terminal. We note that Eq. (21.5.5) agrees with (21.4.3) of the transmitting case.

Equations (21.5.3) and (21.5.4) are special cases of a more general result, which is a consequence of the reciprocity principle (for example, see [34]). Given an incident field on a receiving linear antenna, the induced short-circuit current and open-circuit voltage at its terminals are given by:

$$\boxed{I_{\rm sc}=\frac{1}{V_0}\int_{-h}^{h}E_{\rm in}(z)I(z)\,dz\,,\quad V_{\rm oc}=-\frac{1}{I_0}\int_{-h}^{h}E_{\rm in}(z)I(z)\,dz}\tag{21.5.6}$$

where $I(z)$ is the current generated by V_0 when the antenna is transmitting. Inserting Eq. (21.4.2) into (21.5.6), we can easily derive Eqs. (21.5.3) and (21.5.4). We will use (21.5.6) in Sec. 22.2 to derive the mutual impedance between two antennas.

Finally, we consider case (c) of an arbitrary load impedance Z_L. The current will be continuous across the gap but it does not have to vanish at $z=0$. The voltage difference across the gap will be equal to the voltage drop across the load, that is, $V_L=-Z_LI(0)$. The approximate Hallén equation is now:

$$\bar{Z}I(z)=V(z)=\begin{cases}C_1\cos kz+C_2\sin kz+\dfrac{2E_0}{k\sin\theta}e^{jkz\cos\theta}\,, & z\ge 0\\[2ex] D_1\cos kz+D_2\sin kz+\dfrac{2E_0}{k\sin\theta}e^{jkz\cos\theta}\,, & z\le 0\end{cases}$$

where $D_1=C_1$ because of the continuity of $I(z)$ at $z=0$. The end conditions, $I(h)=I(-h)=0$, give:

$$C_1\cos kh+C_2\sin kh+\frac{2E_0}{k\sin\theta}e^{jkh\cos\theta}=0$$

$$C_1\cos kh-D_2\sin kh+\frac{2E_0}{k\sin\theta}e^{-jkh\cos\theta}=0$$

Moreover, we have the discontinuity condition:

$$V'(0+)-V'(0-)=2kV_L=k(C_2-D_2)\qquad\Rightarrow\qquad V_L=\frac{1}{2}(C_2-D_2)$$

Ohm's law at the load gives:

$$V_L=-Z_LI(0)=-\frac{Z_L}{\bar{Z}}\left(C_1+\frac{2E_0}{k\sin\theta}\right)=\frac{Z_L}{Z_A}\left(C_1+\frac{2E_0}{k\sin\theta}\right)\cot kh$$

where we used Eq. (21.5.5). Solving the above four equations for C_1, C_2, D_2, V_L, we find eventually:

$$V_L=\frac{Z_L}{Z_A+Z_L}\,\frac{2E_0}{k\sin kh}\,\frac{\cos kh-\cos(kh\cos\theta)}{\sin\theta}=\frac{V_{\rm oc}Z_L}{Z_A+Z_L}\tag{21.5.7}$$

This is equivalent to the Thevénin model that we used in Sec. 15.4. The power delivered to the load will be proportional to $|V_L|^2$, which is proportional to the gain pattern of a transmitting dipole, that is,

$$\left|\frac{\cos kh-\cos(kh\cos\theta)}{\sin\theta}\right|^2$$

21.6 King's Three-Term Approximation

To improve the crude sinusoidal approximation of Eq. (21.4.1), we must look more carefully at the properties of the kernel. Separating its real and imaginary parts, we have:

$$\frac{j\eta}{2\pi}G_{\rm app}(z-z')=\frac{j\eta}{2\pi}\,\frac{e^{-jkR}}{R}=\frac{k\eta}{2\pi}\left[\frac{\sin kR}{kR}+j\frac{\cos kR}{kR}\right]$$

For R near zero, the imaginary part becomes very large and we may apply the approximation (21.4.1) to it. But, the real part remains finite at $R=0$. For $kR\le\pi$, which will be guaranteed if $kh\le\pi$, the sinc function can be very well approximated by $\cos(kR/2)\simeq\cos(k|z-z'|/2)$ as can be verified by plotting the two functions. Therefore,

$$\frac{\sin kR}{kR}\simeq\cos(kR/2)\simeq\cos\bigl(k(z-z')/2\bigr)\,,\qquad\text{for}\quad kR\le\pi\tag{21.6.1}$$

Using this approximation for the real part of the kernel, and applying the approximation of Eq. (21.4.1) to its imaginary part, King has shown [4,71] that an improved approximation of the convolution integral is as follows:

$$\frac{j\eta}{2\pi}\int_{-h}^{h}G_{\rm app}(z-z')I(z')\,dz'\simeq R\cos(\frac{kz}{2})+jXI(z)\tag{21.6.2}$$

where R, X are appropriate constants, which are real if $I(z)$ is real. The approximation also assumes that the current is symmetric, $I(z)=I(-z)$. Indeed, we have:

$$\begin{aligned}\frac{j\eta}{2\pi}\int_{-h}^{h}G_{\text{app}}(z-z')I(z')\,dz' &= \frac{k\eta}{2\pi}\int_{-h}^{h}\Big[\cos\big(\frac{k(z-z')}{2}\big)+j\frac{\cos kR}{kR}\Big]I(z')\,dz'\\ &= \frac{k\eta}{2\pi}\int_{-h}^{h}\Big[\cos\big(\frac{kz}{2}\big)\cos\big(\frac{kz'}{2}\big)I(z')+\sin\big(\frac{kz}{2}\big)\sin\big(\frac{kz'}{2}\big)I(z')+j\frac{\cos kR}{kR}I(z')\Big]dz'\end{aligned}$$

The first term is of the form $R\cos(kz/2)$, the second term vanishes because of the assumed even symmetry of $I(z)$, and the third term is of the form $jXI(z)$. It follows that the Hallén equation (21.3.12) can be approximated by:

$$R\cos\big(\frac{kz}{2}\big)+jXI(z)=V(z)=C_1\cos kz+V_0\sin k|z|$$

This shows that the current $I(z)$ is a linear combination of the sinusoidal terms $\sin k|z|$, $\cos kz$, and $\cos(kz/2)$, and leads to King's three-term approximation for the current [4,71], which incorporates the condition $I(h)=0$. There are two alternative forms:

$$I(z)=A_1I_1(z)+A_2I_2(z)+A_3I_3(z)=A_1'I_1'(z)+A_2'I_2'(z)+A_3'I_3'(z) \tag{21.6.3}$$

where the expansion currents are defined by:

$$\boxed{\begin{aligned}I_1(z)&=\sin k|z|-\sin kh\\ I_2(z)&=\cos kz-\cos kh\\ I_3(z)&=\cos(kz/2)-\cos(kh/2)\end{aligned}}\,,\quad \boxed{\begin{aligned}I_1'(z)&=\sin\big(k(h-|z|)\big)\\ I_2'(z)&=\cos kz-\cos kh\\ I_3'(z)&=\cos(kz/2)-\cos(kh/2)\end{aligned}} \tag{21.6.4}$$

Using the trigonometric identity $I_1(z)=I_2'(z)\tan kh-I_1'(z)/\cos kh$, the relationship between the primed and unprimed coefficients is:

$$\begin{bmatrix}A_1'\\ A_2'\end{bmatrix}=\frac{1}{\cos kh}\begin{bmatrix}-1 & 0\\ \sin kh & \cos kh\end{bmatrix}\begin{bmatrix}A_1\\ A_2\end{bmatrix},\quad A_3'=A_3 \tag{21.6.5}$$

The condition number of the transformation matrix is $1/|\cos kh|$, and the transformation breaks down when $\cos kh=0$, that is, when the antenna length $l=2h$ is an odd-multiple of $\lambda/2$. In that case, only the unprimed form may be used. Otherwise, the primed form is preferable because the term $I_1'(z)=\sin\big(k(h-|z|)\big)$ has the conventional standing-wave form. We will work with the unprimed form because it is always possible. The MATLAB function `kingprime` transforms the unprimed coefficients into the primed ones:

```
Aprime = kingprime(L,A);          % converts from unprimed to primed form
```

To determine the expansion coefficients A_1, A_2, A_3, we insert Eq. (21.6.3) into Hallén's equation (21.3.12) and get:

$$A_1V_1(z)+A_2V_2(z)+A_3V_3(z)=V(z)=C_1\cos kz+V_0\sin k|z| \tag{21.6.6}$$

where

$$V_i(z)=\frac{j\eta}{2\pi}\int_{-h}^{h}G_{\text{app}}(z-z')I_i(z')\,dz',\qquad i=1,2,3 \tag{21.6.7}$$

At $z=h$, we have:

$$A_1V_1(h)+A_2V_2(h)+A_3V_3(h)=V(h)=C_1\cos kh+V_0\sin kh \tag{21.6.8}$$

Subtracting Eqs. (21.6.6) and (21.6.8), and defining $V_{di}(z)=V_i(z)-V_i(h)$, we have:

$$A_1V_{d1}(z)+A_2V_{d2}(z)+A_3V_{d3}(z)=C_1(\cos kz-\cos kh)+V_0(\sin k|z|-\sin kh)$$

Using the definition (21.6.4), we can write:

$$A_1V_{d1}(z)+A_2V_{d2}(z)+A_3V_{d3}(z)=C_1I_2(z)+V_0I_1(z) \tag{21.6.9}$$

Introducing the difference kernel $G_d(z-z')=G_{\text{app}}(z-z')-G_{\text{app}}(h-z')$, we have:

$$V_{di}(z)=\frac{j\eta}{2\pi}\int_{-h}^{h}G_d(z-z')I_i(z')\,dz',\qquad i=1,2,3 \tag{21.6.10}$$

The improved approximation (21.6.2) applied to the difference kernel gives:

$$\frac{j\eta}{2\pi}\int_{-h}^{h}G_d(z-z')I(z')\,dz'=R\big(\cos(kz/2)-\cos(kh/2)\big)+jXI(z)=RI_3(z)+jXI(z)$$

Therefore, applying it to the three separate currents $I_1(z), I_2(z), I_3(z)$, we obtain:

$$V_{di}(z)=V_i(z)-V_i(h)=R_iI_3(z)+jX_iI_i(z),\qquad i=1,2,3 \tag{21.6.11}$$

Inserting these approximations in Eq. (21.6.6), we have:

$$\begin{aligned}A_1\big[R_1I_3(z)+jX_1I_1(z)\big]+A_2\big[R_2I_3(z)+jX_2I_2(z)\big]+A_3\big[R_3I_3(z)+jX_3I_3(z)\big]&=\\ =C_1I_2(z)+V_0I_1(z)&\end{aligned}$$

Defining $Z_3=R_3+jX_3$ and matching the coefficients of $I_1(z), I_2(z), I_3(z)$ in the two sides, gives three equations in the four unknowns A_1, A_2, A_3, C_1:

$$jX_1A_1=V_0,\quad jX_2A_2-C_1=0,\quad R_1A_1+R_2A_2+Z_3A_3=0$$

The fourth equation is (21.6.8). Thus, we obtain the linear system:

$$\begin{bmatrix}jX_1 & 0 & 0 & 0\\ 0 & jX_2 & 0 & -1\\ R_1 & R_2 & Z_3 & 0\\ V_1(h) & V_2(h) & V_3(h) & -\cos kh\end{bmatrix}\begin{bmatrix}A_1\\ A_2\\ A_3\\ C_1\end{bmatrix}=\begin{bmatrix}V_0\\ 0\\ 0\\ V_0\sin kh\end{bmatrix} \tag{21.6.12}$$

The matrix elements can be determined by evaluating the defining approximations (21.6.11) at z-points at which the currents $I_i(z)$ take on their maximum values. For $I_1(z)$, the maximum occurs at $z_1=0$ if $h\le\lambda/4$ and at $z_1=h-\lambda/4$ if $\lambda/4\le h\le 5\lambda/8$.

For $I_2(z)$ and $I_3(z)$, the maxima occur at $z = 0$. Thus, the defining equations for the matrix elements are:

$$\boxed{\begin{aligned} V_{d1}(z_1) &= V_1(z_1) - V_1(h) = R_1 I_3(z_1) + jX_1 I_1(z_1) \\ V_{d2}(0) &= V_2(0) - V_2(h) = R_2 I_3(0) + jX_2 I_2(0) \\ V_{d3}(0) &= V_3(0) - V_3(h) = Z_3 I_3(0) \end{aligned}} \tag{21.6.13}$$

The coefficients R_1, X_1, R_2, X_2 are obtained by extracting the real and imaginary parts of these expressions. The left-hand sides can be computed by direct numerical integration of the definitions (21.6.7). The expected range of applicability of the 3-term approximation is for antenna lengths $l \le 1.25\lambda$ (see [4,71].) However, it works well even for longer lengths.

The MATLAB function `king` implements the design equations (21.6.12) and (21.6.13). It has usage:

```
A = king(L,a);                % King's 3-term sinusoidal approximation
```

where L, a are the antenna length and its radius in units of λ and the output $\mathbf{A}$ is the column vector of the coefficients A_i. If the length is an odd-multiple of $\lambda/2$, then $\mathbf{A} = [A_1, A_2, A_3]^T$, otherwise, $\mathbf{A} = [A_1', A_2', A_3']^T$.

The numerical integrations are done with a 32-point Gauss-Legendre quadrature integration routine implemented with the function `quadr`, which provides the appropriate weights and evaluation points for the integration.

Example 21.6.1: Fig. 21.6.1 compares the three-term approximation to the standard sinusoidal approximation, $I(z) = \sin\big(k(h - |z|)\big)$, and to the exact numerical solution of Hallén's equation for the two cases of $l = \lambda$ and $l = 1.5\lambda$. The antenna radius was $a = 0.005\lambda$.

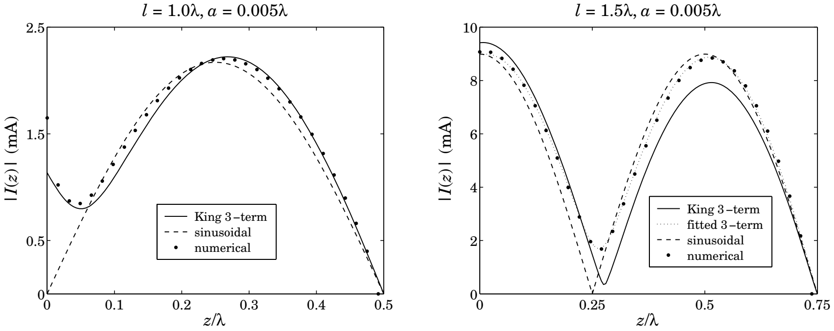


Fig. 21.6.1 Three-term approximation for $l = \lambda$ and $l = 1.5\lambda$.

In the full-wavelength case, the sinusoidal approximation has $I(0) = 0$, which would imply infinite antenna impedance. The three-term approximation gives a nonzero value for $I(0)$. The computed three-term coefficients are in the two cases:

$$\begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = 10^{-3} \begin{bmatrix} -2.6035j \\ 0.2737 + 0.2779j \\ 0.2666 + 0.2376j \end{bmatrix}, \quad \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = 10^{-3} \begin{bmatrix} -2.1403j \\ 7.7886 - 3.6840j \\ 0.8688 + 2.4546j \end{bmatrix}$$

We used the unprimed representation for both cases (the primed one coincides with the unprimed one for the case $l = \lambda$ because $\cos kh = -1$ and the transformation matrix (21.6.5) becomes the identity matrix.) The graphs were generated by the following example code (for the $l = 1.5\lambda$ case):

```
L = 1.5;  h = L/2; a = 0.005;      % length and radius
k = 2*pi;                          % wavenumber in units of λ = 1

M = 30;                            % number of cells is 2 ∗ M + 1
[In,zn] = hdelta(L,a,M,'e');       % numerical solution of Hallén equation with exact kernel
In = In(M+1:end);                  % keep only upper half of the values
zn = zn(M+1:end);

A = king(L,a);                     % King's three-term approximation
z = 0:h/150:h;                     % evaluation points on upper half
Ik = abs(kingeval(L,A,z));         % evaluate King's three-term current

B = kingfit(L,In,zn,1);            % fit one-term sinusoidal current
I1 = abs(kingeval(L,B,z));         % evaluate one-term sinusoidal current

C = kingfit(L,In,zn,3);            % fit three-term current to the numerical values
I3 = abs(kingeval(L,C,z));         % evaluate fitted three-term current

plot(z,Ik,'-', z,I3,':', z,I1,'--', zn,abs(In), '.');
```

The currents $I_1(z)$ and $I_3(z)$ represent the one-term and three-term fits to the numerical samples I_n at the points z_n, as described below. ☐

As is evident from the above example, King's three-term approximation does not work particularly well for larger antenna lengths (about $l > 1.25\lambda$). This can be attributed to the crude approximation of computing the coefficients A_i by matching the defining currents only at one point along the antenna (at the current maxima).

It turns out, however, that the three-term approximation is very accurate if fitted to the "exact" current as computed by solving Hallén's equation numerically, with a range of applicability of up to about $l = 2\lambda$. With a 4-term fit, the range increases to $l = 3\lambda$.

Typically, numerical methods generate a set of N current values I_n at N points z_n, $n = 1, 2, \dots, N$, along the antenna. These values can be fitted to a three-term expression of the form of Eq. (21.6.3) using the least-squares criterion:

$$\mathcal{J} = \sum_{n=1}^{N} \big|I_s(z_n) - I_n\big|^2 = \min, \quad \text{where} \quad I_s(z) = \sum_{i=1}^{3} A_i I_i(z) \tag{21.6.14}$$

where $\mathcal{J}$ is minimized with respect to the three coefficients A_1, A_2, A_3. This is equivalent to finding the least-squares solution of the overdetermined $N\times 3$ linear system of equations (assuming $N > 3$):

$$\begin{bmatrix} I_1(z_1) & I_2(z_1) & I_3(z_1) \\ \vdots & \vdots & \vdots \\ I_1(z_n) & I_2(z_n) & I_3(z_n) \\ \vdots & \vdots & \vdots \\ I_1(z_N) & I_2(z_N) & I_3(z_N) \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ I_n \\ \vdots \\ I_N \end{bmatrix} \tag{21.6.15}$$

Writing this system in the compact matrix form $S\mathbf{A} = \mathbf{I}$, its MATLAB solution is obtained by the backslash operation: $\mathbf{A} = S\backslash\mathbf{I}$. More generally, one may perform the fit to $p = 1, 2, 3, 4$ sinusoidal terms, that is,

$$I_s(z) = \sum_{i=1}^{p} A_i I_i(z) \tag{21.6.16}$$

For $p = 1, 2, 3$, the basis currents $I_i(z)$ are as in Eq. (21.6.4). For $p = 1$, the basis is always defined as $I_1(z) = \sin(kh - k|z|)$. For $p = 4$, the first two basis currents, $I_1(z), I_2(z)$, are as in (21.6.4), and the last two are:

$$\begin{aligned} I_3(z) &= \cos(kz/4) - \cos(kh/4) \\ I_4(z) &= \cos(3kz/4) - \cos(3kh/4) \end{aligned} \tag{21.6.17}$$

The MATLAB function `kingfit` solves the system of equations (21.6.15), or its more general version, and returns the coefficients A_i. It has the following usage, where p is the desired number of terms:

```
A = kingfit(L,In,zn,p);        % p-term fit to sinusoidal currents
```

The function `kingeval` evaluates the p-term approximation (21.6.16) at a given number of z-points:

```
I = kingeval(L,A,z);        % evaluate p-term expression I(z) at the points z
```

where the number of terms p is determined from the number of coefficients A_i. The right graph of Fig. 21.6.1 compares King's and the least-squares three-term approximations.

The four-term approximation is justified as follows. The three-term case was based on the approximation $\sin kR/kR \simeq \cos(kR/2)$. To improve it, we consider the identity:

$$\frac{\sin kR}{kR} = \frac{\sin(kR/2)}{kR/2}\cos(kR/2) = \frac{\sin(kR/4)}{kR/4}\cos(kR/4)\cos(kR/2)$$

The three-term case is obtained by replacing $\sin(kR/2)/(kR/2) \simeq 1$, which is approximately valid for $R \le \lambda/2$. A better approximation is obtained from the second identity by setting $\sin(kR/2)/(kR/2) \simeq 1$. This results in the approximation:

$$\frac{\sin kR}{kR} \simeq \cos(kR/4)\cos(kR/2) = \frac{1}{2}\big[\cos(kR/4) + \cos(3kR/4)\big] \tag{21.6.18}$$

which is well satisfied up to $R \le 3\lambda/2$. Using the same arguments that led to Eq. (21.6.2), we now obtain the approximation:

$$\frac{j\eta}{2\pi}\int_{-h}^{h} G_{\text{app}}(z - z')I(z')dz' \simeq R\cos\Big(\frac{kz}{4}\Big) + R'\cos\Big(\frac{3kz}{4}\Big) + jXI(z) \tag{21.6.19}$$

where R, R', X are appropriate constants. Thus, Hallén's equation (21.3.12) can be approximated as:

$$R\cos\Big(\frac{kz}{4}\Big) + R'\cos\Big(\frac{3kz}{4}\Big) + jXI(z) = V(z) = C_1\cos kz + V_0\sin k|z|$$

which implies that $I(z)$ can be written as the sum of four sinusoidal currents, $I_1(z), I_2(z)$, given by Eq. (21.6.3), and $I_3(z), I_4(z)$, given by (21.6.17).

Fig. 21.6.2 compares the three-term and four-term fits for the two antenna lengths $l = \lambda$ and $l = 3\lambda$. For the $l = \lambda$ case, the two fits are virtually indistinguishable. The antenna radius was $a = 0.005\lambda$ and the "exact" numerical solution was computed using the exact kernel with $2M + 1 = 101$ segments. The graphs can be generated by the following example code:

```
L=3; a=0.005; M=50;

[Ie,z] = hdelta(L,a,M,'e');        % solve Hallén equation with exact kernel and delta-gap input

A = kingfit(L,Ie,z,3); I3 = kingeval(L,A,z);
B = kingfit(L,Ie,z,4); I4 = kingeval(L,B,z);

plot(z,real(Ie),'.', z,real(I4),'-', z,real(I3),'--');
```

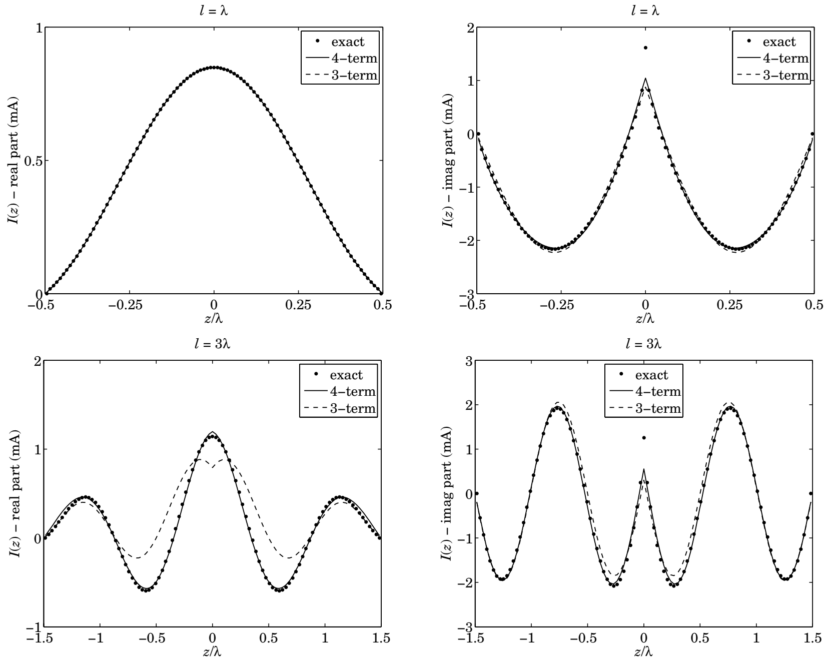


Fig. 21.6.2 Three- and four-term approximations for $l = \lambda$ and $l = 3\lambda$.

We will look at further examples later on. The main advantage of such fits is that they provide simple analytical expressions for the current, which can be used in turn to compute the radiation pattern. We saw in Eq. (16.1.7) that the radiation intensity of a linear antenna is given by

$$U(\theta)=\frac{\eta k^2}{32\pi^2}\,|F_z(\theta)|^2\sin^2\theta$$

where $F_z(\theta)$ is the z-component of the radiation vector:

$$F_z(\theta)=\int_{-h}^{h}I(z)e^{jkz\cos\theta}dz$$

For the p-term current given by Eq. (21.6.16), we have:

$$F_z(\theta)=\sum_{i=1}^{p}A_i\int_{-h}^{h}I_i(z)e^{jkz\cos\theta}dz=\sum_{i=1}^{p}A_iF_i(\theta)\tag{21.6.20}$$

The individual radiation vectors $F_i(\theta)$ are given by closed-form expressions as follows. For $I_1(z)$ and $I_1'(z)$, we have:

$$\begin{aligned}F_1(\theta)&=\int_{-h}^{h}(\sin k|z|-\sin kh)e^{jkz\cos\theta}dz=\\&=\frac{1}{k}\,\frac{(1-\cos(kh\cos\theta)\cos kh)\cos\theta-\sin(kh\cos\theta)\sin kh}{\cos\theta\sin^2\theta}\\F_1'(\theta)&=\int_{-h}^{h}\sin(kh-k|z|)e^{jkz\cos\theta}dz=\frac{2}{k}\,\frac{\cos(kh\cos\theta)-\cos kh}{\sin^2\theta}\end{aligned}\tag{21.6.21}$$

The rest of the radiation vectors are obtained from the following integral, with the parameter values $\alpha=1,\,1/2,\,1/4,\,3/4$:

$$\begin{aligned}&\int_{-h}^{h}\big[\cos(k\alpha z)-\cos(k\alpha h)\big]e^{jkz\cos\theta}dz=\\&=\frac{\alpha}{k}\,\frac{(\alpha+\cos\theta)\sin\big(kh(\alpha-\cos\theta)\big)-(\alpha-\cos\theta)\sin\big(kh(\alpha+\cos\theta)\big)}{\cos\theta\,(\alpha^2-\cos^2\theta)}\end{aligned}\tag{21.6.22}$$

21.7 Evaluation of the Exact Kernel

Numerical methods for Hallén's and Pocklington's equations require the numerical evaluation (and integration) of the exact or approximate kernel. A sample of such numerical methods is given in Refs. [1207–1266].

The evaluation of the approximate kernel is straightforward. The exact kernel requires a more careful treatment because of its singularity at $z=0$. Here, we follow [1260] and express the exact kernel in terms of elliptic functions and discuss its numerical evaluation. The exact kernel was defined in Eq. (21.1.2):

$$G(z,\rho)=\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-jkR}}{R}\,d\phi'\,,\quad R=\sqrt{z^2+\rho^2+a^2-2\rho a\cos\phi'}\tag{21.7.1}$$

The distance R may be written in the alternative forms:

$$\begin{aligned}R&=\sqrt{z^2+(\rho+a)^2-2\rho a(1+\cos\phi')}\\&=\sqrt{z^2+(\rho+a)^2-4\rho a\cos^2(\phi'/2)}\\&=\sqrt{z^2+(\rho+a)^2-4\rho a\sin^2\theta}\\&=R_{\max}\sqrt{1-\kappa^2\sin^2\theta}\end{aligned}\tag{21.7.2}$$

where we defined:

$$R_{\max}=\sqrt{z^2+(\rho+a)^2}\,,\quad \kappa=\frac{2\sqrt{a\rho}}{R_{\max}}=\frac{2\sqrt{a\rho}}{\sqrt{z^2+(\rho+a)^2}}\tag{21.7.3}$$

and made the change of variables $\phi'=\pi+2\theta$. Under this change, the integration range $[0,2\pi]$ in ϕ' maps onto $[-\pi/2,\pi/2]$ in θ, and because R is even in θ, that range can be further reduced to $[0,\pi/2]$, resulting into the expression for the kernel:

$$G(z,\rho)=\frac{2}{\pi}\int_0^{\pi/2}\frac{e^{-jkR}}{R}\,d\theta=\frac{2}{\pi R_{\max}}\int_0^{\pi/2}\frac{e^{-jkR_{\max}\sqrt{1-\kappa^2\sin^2\theta}}}{\sqrt{1-\kappa^2\sin^2\theta}}\,d\theta\tag{21.7.4}$$

where $R_{\max}$ represents the maximum value of R as θ varies. The approximate kernel corresponds to the limit $a=0$ or $\kappa=0$. The connection to elliptic functions comes about as follows [1267–1271]. The change of variables,

$$u=\int_0^{\theta}\frac{d\alpha}{\sqrt{1-\kappa^2\sin^2\alpha}}\quad\Rightarrow\quad du=\frac{d\theta}{\sqrt{1-\kappa^2\sin^2\theta}}\tag{21.7.5}$$

defines θ indirectly as a function of u. The Jacobian elliptic functions $\mathrm{sn}(u,\kappa)$ and $\mathrm{dn}(u,k)$ are then defined by

$$\begin{aligned}\mathrm{sn}(u,\kappa)&=\sin\theta\\\mathrm{dn}(u,k)&=\sqrt{1-\kappa^2\mathrm{sn}^2(u,\kappa)}=\sqrt{1-\kappa^2\sin^2\theta}\end{aligned}\tag{21.7.6}$$

where κ is referred to as the elliptic *modulus*. The complete elliptic integrals of the first and second kinds are given by:

$$K(\kappa)=\int_0^{\pi/2}\frac{d\theta}{\sqrt{1-\kappa^2\sin^2\theta}}\,,\quad E(\kappa)=\int_0^{\pi/2}\sqrt{1-\kappa^2\sin^2\theta}\,d\theta\tag{21.7.7}$$

Thus, when $\theta=\pi/2$, then $u=K(\kappa)$. With these definitions, Eq. (21.7.4) can be written as:

$$G(z,\rho)=\frac{2}{\pi R_{\max}}\int_0^{K(\kappa)}e^{-jkR_{\max}\mathrm{dn}(u,\kappa)}\,du\tag{21.7.8}$$

Changing variables from u to $uK(\kappa)$, we may write:

$$G(z,\rho)=\frac{2K(\kappa)}{\pi R_{\max}}\int_0^{1}e^{-jkR_{\max}\mathrm{dn}(uK,\kappa)}\,du\tag{21.7.9}$$

For points on the surface of the antenna wire $(\rho = a)$, the kernel and the quantities $R_{\max}$ and κ simplify into:

$$\boxed{G(z) = \frac{2}{\pi}\int_0^{\pi/2}\frac{e^{-jkR}}{R}\,d\theta = \frac{2K(\kappa)}{\pi R_{\max}}\int_0^1 e^{-jkR_{\max}\,\mathrm{dn}(uK,\kappa)}\,du} \quad \text{(exact kernel)} \tag{21.7.10}$$

with $R = \sqrt{z^2+4a^2-4a^2\sin^2\theta} = R_{\max}\sqrt{1-\kappa^2\sin^2\theta}$ and

$$R_{\max} = \sqrt{z^2+4a^2}\,, \qquad \kappa = \frac{2a}{R_{\max}} = \frac{2a}{\sqrt{z^2+4a^2}} \tag{21.7.11}$$

As u varies over the interval $0 \le u \le 1$, the quantity $\mathrm{dn}(uK,\kappa)$ stays bounded, varying over the range:

$$\kappa' \le \mathrm{dn}(uK,\kappa) \le 1 \tag{21.7.12}$$

where we introduced the *complementary* modulus:

$$\kappa' = \sqrt{1-\kappa^2} = \frac{|z|}{\sqrt{z^2+4a^2}} = \frac{|z|}{R_{\max}} \tag{21.7.13}$$

Therefore, the integral in Eq. (21.7.10) remains bounded and less than one in magnitude for all values of z. On the other hand, the factor $K(\kappa)$ incorporates the logarithmic singularity at $z = 0$. Indeed, as $z \to 0$, the moduli κ and κ' tend to 1 and 0, respectively, and $K(\kappa)$ behaves as $\ln(4/\kappa')$ [1270]:

$$K(\kappa) \simeq \ln\left(\frac{4}{\kappa'}\right) \simeq \ln\left(\frac{4R_{\max}}{|z|}\right) \simeq \ln\left(\frac{8a}{|z|}\right), \quad \text{as } z \to 0 \tag{21.7.14}$$

where we replaced $R_{\max} \simeq 2a$ as $z \to 0$. Thus, the kernel behaves like

$$G(z) \simeq \frac{1}{\pi a}\ln\left(\frac{8a}{|z|}\right), \quad \text{as } z \to 0 \tag{21.7.15}$$

The MATLAB function `kernel` implements Eq. (21.7.10) to compute $G(z)$ at any vector of z points. For smaller values of z, it uses the asymptotic form (21.7.15). It has usage:

```
G = kernel(z,a,'e');        % exact kernel
G = kernel(z,a,'a');        % approximate kernel
```

It employs the following set of MATLAB functions for the evaluation of the complete elliptic integrals and the function $\mathrm{dn}(uK,\kappa)$:

```
K = ellipK(k);       % elliptic integral K(κ) at a vector of κ's
E = ellipE(k);       % elliptic integral E(κ) at a vector of κ's
v = landenv(k);      % Landen transformations of a vector of κ's
w = snv(u,k);        % sn(uK,κ) function at a vector of u's and a vector of κ's
w = dnv(u,k);        % dn(uK,κ) function at a vector of u's and a vector of κ's
```

These are based on a set of similar functions developed for the implementation of elliptic filters [1272–1275] that were modified here to handle a vector of moduli κ arising from a vector of z points. Using these functions, the integral in Eq. (21.7.10) is implemented with a 32-point Gauss-Legendre integration over the interval $0 \le u \le 1$. Let w_i, u_i, $i = 1,2,\dots,32$, denote the weights and evaluation points obtained by calling the quadrature function `quadr`:

```
[w,u] = quadr(0,1,32);       % 32-point Gauss-Legendre integration over the interval [0,1]
```

Then, Eq. (21.7.10) can be evaluated by

$$G(z) = \frac{2K(\kappa)}{\pi R_{\max}}\sum_{i=1}^{32} w_i\, e^{-jkR_{\max}\,\mathrm{dn}(u_i K,\kappa)} \tag{21.7.16}$$

The function `kernel` has an additional input parameter, `method`,

```
G = kernel(z,a,'e',method);        % exact kernel
```

that allows one to select faster but somewhat less accurate methods of computing the kernel. The method of Eq. (21.7.16) is selected with `method` = 3. The integral in (21.7.10) can be expanded approximately as follows [1260]:

$$\begin{aligned}
J(\kappa) &= K\int_0^1 e^{-jkR_{\max}\,\mathrm{dn}(uK,\kappa)}\,du = \int_0^K e^{-jkR_{\max}\,\mathrm{dn}(u,\kappa)}\,du \\
&= e^{-jkR_{\max}}\int_0^K e^{-jkR_{\max}\left(\mathrm{dn}(u,\kappa)-1\right)}\,du \\
&\simeq e^{-jkR_{\max}}\int_0^K\left[1 - jkR_{\max}\left(\mathrm{dn}(u,\kappa)-1\right) + \frac{(-jkR_{\max})^2}{2}\left(\mathrm{dn}(u,\kappa)-1\right)^2\right]du
\end{aligned}$$

Using the definitions (21.7.5)–(21.7.7), we find:

$$\int_0^K\left(\mathrm{dn}(u,\kappa)-1\right)du = \frac{\pi}{2} - K\,, \qquad \int_0^K\left(\mathrm{dn}(u,\kappa)-1\right)^2 du = K + E - \pi$$

Thus, $J(\kappa)$ can be written approximately as

$$J(\kappa) = \int_0^K e^{-jkR_{\max}\,\mathrm{dn}(u,\kappa)} \simeq e^{-jkR_{\max}}\left[K + jkR_{\max}\left(K - \frac{\pi}{2}\right) + \frac{(jkR_{\max})^2}{2}(K + E - \pi)\right]$$

This leads to the following approximations for the kernel $G(z)$. If only the linear term in $(jkR_{\max})$ is kept, then

$$G(z) = \frac{2e^{-jkR_{\max}}}{\pi R_{\max}}\left[K + jkR_{\max}\left(K - \frac{\pi}{2}\right)\right] \tag{21.7.17}$$

and, if both the linear and the quadratic terms are kept,

$$G(z) = \frac{2e^{-jkR_{\max}}}{\pi R_{\max}}\left[K + jkR_{\max}\left(K - \frac{\pi}{2}\right) + \frac{(jkR_{\max})^2}{2}(K + E - \pi)\right] \tag{21.7.18}$$

Eqs. (21.7.17) and (21.7.18) are selected with the `method` = 1, 2, respectively, and provide faster alternatives to the slower but more accurate method of Eq. (21.7.16).

Because $\kappa^2 = 1-\kappa'^2$, floating point accuracy limits the values of κ'^2 to be greater than about the machine epsilon, that is, $\kappa' > \sqrt{\epsilon}$, which for MATLAB gives $\epsilon = 2.22\times10^{-16}$ and $\kappa' > \sqrt{\epsilon} = 1.49\times10^{-8}$. Since for small z we have $\kappa' = z/2a$, this limitation translates to a minimum value of z below which the elliptic function calculations cannot be used and one must use the asymptotic form (21.7.15):

$$\frac{z_{\min}}{2a} = \kappa' = \sqrt{\epsilon} \quad \Rightarrow \quad z_{\min} = (2.98\times10^{-8})\,a \tag{21.7.19}$$

An alternative computation method, which will also help refine the asymptotic form (21.7.15), is based on a straightforward series expansion of the integral in (21.7.10):

$$\begin{aligned} J(\kappa) &= K\int_0^1 e^{-jkR_{\max}\,\mathrm{dn}(uK,\kappa)}\,du = \int_0^K e^{-jkR_{\max}\,\mathrm{dn}(u,\kappa)}\,du \\ &= \sum_{m=0}^{\infty}\frac{(-jkR_{\max})^m}{m!}\int_0^K \mathrm{dn}^m(u,\kappa)\,du \end{aligned}$$

Defining the integrals,

$$J_m(\kappa)=\int_0^K \mathrm{dn}^m(u,\kappa)\,du = \int_0^{\pi/2}\left(\sqrt{1-\kappa^2\sin^2\theta}\right)^{m-1} d\theta\,, \quad m\ge 0 \tag{21.7.20}$$

we have:

$$J(\kappa)=\sum_{m=0}^{\infty}\frac{(-jkR_{\max})^m}{m!}\,J_m(\kappa) \tag{21.7.21}$$

The first few of these are:

$$\begin{aligned} J_0(\kappa) &= K(\kappa) \\ J_1(\kappa) &= \frac{\pi}{2} \\ J_2(\kappa) &= E(\kappa) \\ J_3(\kappa) &= \frac{\pi}{4}(1+\kappa'^2) \\ J_4(\kappa) &= \frac{1}{3}\big[2(1+\kappa'^2)E(\kappa)-\kappa'^2K(\kappa)\big] \end{aligned} \tag{21.7.22}$$

where $\kappa'^2 = 1-\kappa^2$. The rest can be computed from the following recursion [1270]:

$$J_{m+1}(\kappa)=\frac{(m-1)(1+\kappa'^2)J_{m-1}(\kappa)-(m-2)\kappa'^2J_{m-3}(\kappa)}{m}\,, \quad m\ge 4 \tag{21.7.23}$$

Separating the $m=0$ term from the rest, the kernel can be written in the form:

$$G(z)=\frac{2}{\pi R_{\max}}\big(K(\kappa)+C(\kappa)\big)\,, \quad C(\kappa)=\sum_{m=1}^{\infty}\frac{(-jkR_{\max})^m}{m!}\,J_m(\kappa) \tag{21.7.24}$$

In the limit $\kappa\to 1$, the quantities $J_m(\kappa)$ have a finite limit, with the exception of $J_0(\kappa)$, which diverges as $J_0(\kappa)=K(\kappa)=\ln(4/\kappa')$. For example, the term $\kappa'^2K(\kappa)$ in $J_4(\kappa)$ converges to zero:

$$\lim_{\kappa\to 1}\kappa'^2K(\kappa)=\lim_{\kappa'\to 0}k'^2\ln\left(\frac{4}{\kappa'}\right)=0$$

In this limit, the integrals in (21.7.20) can be done in closed form and expressed in terms of the gamma function [1267]:

$$J_m(1)=\int_0^{\pi/2}(\cos\theta)^{m-1}d\theta = \frac{\pi}{2}\,\frac{\Gamma\left(\dfrac{m}{2}\right)}{\Gamma\left(\dfrac{m+1}{2}\right)}\,, \quad m\ge 1 \tag{21.7.25}$$

Therefore, the term $C(\kappa)$ in (21.7.24) also has a finite limit given by:

$$C(1)=\sum_{m=1}^{\infty}\frac{(-2jka)^m}{m!}\,J_m(1) \tag{21.7.26}$$

where we replaced $R_{\max}=2a$. Then, the asymptotic form (21.7.15) of the kernel may be modified by adding the constant $C(1)$:

$$G(z)=\frac{1}{\pi a}\left[\ln\left(\frac{8a}{|z|}\right)+C(1)\right]\,, \quad \text{as } z\to 0 \tag{21.7.27}$$

In the function `kernel`, we use this approximation for $|z|<z_{\min}$, where the series for $C(1)$ converges very fast requiring about 4–5 terms for typical values of the radius, such as $0.001\lambda\le a\le 0.01\lambda$.

For $|z|\ge z_{\min}$, one may use the series expansion (21.7.24). The smaller the z, the smaller the number of terms required for convergence of the sum. For example, for the case $a=0.005\lambda$ and for some typical values of z, the number of terms were:

z/λ	3	2	1	0.5	0.05	0.005	0.0005
M	76	58	38	22	10	7	7

The method (21.7.24) is selected by the value `method = 4` in the function `kernel`. We have included all four methods described by Eqs. (21.7.16)–(21.7.18), and (21.7.24) as options in `kernel`. The default method is that of Eq. (21.7.18) selected with `method = 2`.

21.8 Method of Moments

The method of moments (MoM) refers to a family of numerical methods for solving integral equations [1213–1219]. We summarize the method in the context of solving Hallén's equations for a delta-gap input, and later on, we apply it to the cases of arbitrary incident fields and Pocklington's equation. For an antenna of length l, half-length $h=l/2$, and radius a, Hallén's equation reads:

$$\frac{j\eta}{2\pi}\int_{-h}^{h}G(z-z')I(z')dz' = V(z)=C_1\cos kz+V_0\sin k|z| \tag{21.8.1}$$

where $G(z-z')$ is the exact or the approximate kernel. The antenna is divided into $N=2M+1$ segments of width $\Delta=l/N=2h/(2M+1)$, as shown in Fig. 21.8.1, with centers at the positions:

$$z_m = m\Delta\,, \quad -M\le m\le M \tag{21.8.2}$$

and the current is expanded into a sum of basis functions:

$$I(z')=\sum_{m=-M}^{M}I_m B(z'-z_m) \tag{21.8.3}$$

where $B(z'-z_m)$ are localized functions centered on the mth segment. For example, in the case of pulse basis functions shown in Fig. 21.8.1, we have:

$$B(z'-z_m)=\begin{cases} 1, & \text{if } \ |z'-z_m|\le \frac{1}{2}\Delta \\ 0, & \text{otherwise}\end{cases} \tag{21.8.4}$$

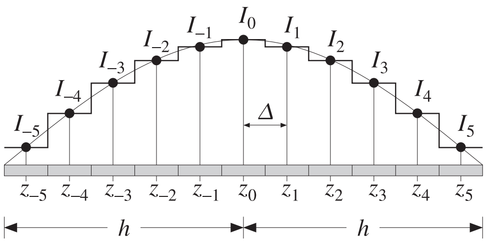


Fig. 21.8.1 Pulse-function segments along an antenna, with $N = 11$, $M = 5$.

Other basis functions are possible such as triangular, sinusoidal, or even plain delta-functions, and we will consider them in succeeding sections. Because of the localized nature of the basis, the expansion (21.8.3) is referred to as a *sub-domain* expansion. Alternatively, *entire-domain* basis functions can be used that are defined over the entire length l of the antenna. Substitution of $I(z')$ into the Hallén equation gives:

$$\frac{j\eta}{2\pi}\sum_{m=-M}^{M} I_m \int_{-h}^{h} G(z-z')B(z'-z_m)dz' = V(z) \tag{21.8.5}$$

The localized nature of $B(z'-z_m)$ restricts this integral to be over the mth segment. Next, a *local weighted average* is formed about each point $z_n = n\Delta$ by using another local weighting (or testing) function $W(z-z_n)$, which is centered on z_n:

$$\frac{j\eta}{2\pi}\sum_{m=-M}^{M} I_m \int_{-h}^{h}\int_{-h}^{h} W(z-z_n)Z(z-z')B(z'-z_m)\,dz\,dz' = \int_{-h}^{h} W(z-z_n)V(z)\,dz$$

This may be written in the $N\times N$ matrix form:

$$\boxed{\sum_{m=-M}^{M} Z_{nm}I_m = v_n\,, \quad -M \le n \le M} \tag{21.8.6}$$

where we defined

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h}\int_{-h}^{h} W(z-z_n)G(z-z')B(z'-z_m)\,dz\,dz' \\ v_n &= \int_{-h}^{h} W(z-z_n)V(z)\,dz \end{aligned} \tag{21.8.7}$$

Two popular choices for the weighting function are the Galerkin and the point-matching or collocation choices. In the Galerkin method the weighting function is taken to be the same as the basis function, and in the point-matching case, it is a delta function:

$$\begin{aligned} W(z-z_n) &= \delta(z-z_n) \qquad \text{(point-matching)} \\ W(z-z_n) &= B(z-z_n) \qquad \text{(Galerkin)} \end{aligned} \tag{21.8.8}$$

Thus, in the *point-matching method*, Eqs. (21.8.7) reduce to:

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n-z')B(z'-z_m)\,dz' \\ v_n &= V(z_n) = C_1\cos kz_n + V_0\sin k|z_n| \end{aligned} \tag{21.8.9}$$

Similarly, in the *Galerkin method*, we have:

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h}\int_{-h}^{h} B(z-z_n)G(z-z')B(z'-z_m)\,dz\,dz' \\ v_n &= \int_{-h}^{h} B(z-z_n)V(z)\,dz \end{aligned} \tag{21.8.10}$$

In succeeding sections, we will consider the following cases and discuss how to compute the quantities Z_{nm} and v_n:

- Delta-function basis with point matching
- Pulse-function basis with point matching
- Pulse-function basis with Galerkin weighting
- Triangular basis with point matching
- Sinusoidal spline basis (NEC basis) with point matching
- Method of moments for Hallén's equation with arbitrary incident field
- Method of moments for Pocklington's equation

We will also consider the following issues that have been discussed extensively in the literature regarding the existence of solutions of Hallén's equation (21.8.1) with delta-gap input [1244–1256]:

1. The approximate kernel is non-singular at $z = 0$. Yet, the numerical solution of Hallén's equation using the approximate kernel does not converge and becomes unusable for increasing number of segments N and/or for increasing radius a, whereas the solution based on the exact kernel does converge. However, for moderate values of N, one does get useful results from the approximate kernel.

2. In fact, the approximate-kernel Hallén equation for a delta-gap input does not, strictly speaking, have a solution, whereas the one with the exact kernel does.

3. The input impedance of the antenna, $Z_0 = V_0/I(0)$, for the delta-gap case does not converge to a constant value for the approximate kernel as N increases, but it does so for the exact kernel. Generally, numerical methods get the resistive part of Z_0 fairly accurately, but have a hard time for the reactive part.

4. The solution $I(z)$ for the exact kernel in the delta-gap case has a logarithmic singularity at $z = 0$ of the form:

$$I(z) \simeq -j\,\frac{4kaV_0}{\eta}\,\ln\bigl(k|z|\bigr)\,, \quad z \simeq 0 \tag{21.8.11}$$

Therefore, one may wonder if the numerical solutions have any use. However, this logarithmic singularity is confined to a very narrow range around $z = 0$ and for all other values of z, the exact-kernel solution is accurate and useful.

5. King's empirical three-term approximation for the current is also very accurate (except in the immediate vicinity of the logarithmic singularity at $z = 0$), if fitted to the exact-kernel solution.

21.9 Delta-Function Basis

Here, we discuss numerical solutions of (21.8.1) using a delta-function basis with point matching. The basis functions are defined by:

$$B(z'-z_m)=\delta(z'-z_m)\Delta \tag{21.9.1}$$

Fig. 21.9.1 depicts the delta functions as narrow pulses of width δ. The factor Δ is needed to give $I(z')$ the right dimensions.

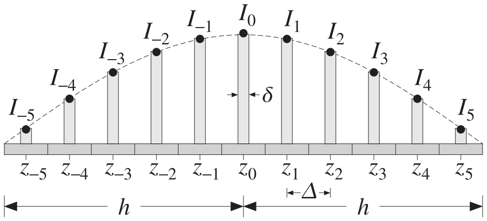


Fig. 21.9.1 Delta-function segments along an antenna, with $N = 11$, $M = 5$.

Inserting this basis into Eq. (21.8.9), we find

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi} G(z_n - z_m)\Delta\,, \quad -M\le n,m\le M \\ v_n &= C_1\cos kz_n + V_0\sin k|z_n|\,, \quad -M\le n\le M \end{aligned} \tag{21.9.2}$$

Because $z_n - z_m = (n-m)\Delta$, the diagonal entries Z_{nn} correspond to the evaluation of the kernel $G(z)$ at $z = 0$. For the approximate kernel, $G(0)$ is finite (corresponding to the small but finite radius $R = a$). But for the exact kernel, $G(0)$ is infinite because of its logarithmic singularity at $z = 0$. A reasonable way of modifying Z_{nn} is to replace the infinitely thin delta-function by a finite-width pulse:

$$B(z)=\delta(z)\Delta \to \frac{u(z+\delta/2)-u(z-\delta/2)}{\delta}\,\Delta$$

where $u(z)$ is the unit-step. Then, Eq. (21.9.3) gives in the case $n = m$:

$$Z_{nn} = \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n - z')B(z'-z_n)\,dz' = \frac{\Delta}{\delta}\,\frac{j\eta}{2\pi}\int_{-\delta/2}^{\delta/2} G(z)\,dz$$

where we used the even-ness of $G(z)$. The best choice for δ which yields results comparable to the other bases is the value $\delta = \Delta$, giving:

$$Z_{nn} = \frac{j\eta}{2\pi}\int_{-\Delta/2}^{\Delta/2} G(z)\,dz\,, \quad -M\le n\le M \tag{21.9.3}$$

We evaluate this integral numerically using Gauss-Legendre integration, However, one could also use quadrature rules [1234,1235] that are more appropriate for the logarithmic singularity of $G(z)$. The above method of regularizing the impedance matrix by local averaging about the singularity region is akin to alternative discretizations based on locally-corrected Nyström methods that replace the integral equation by a quadrature integration rule, such as Gauss-Legendre [1240–1243].

With the definitions (21.9.2) and (21.9.3), the matrix equation (21.8.6) may be written in the compact form:

$$\mathcal{Z}\boldsymbol{I} = \boldsymbol{v} = C_1\mathbf{c} + V_0\mathbf{s} \tag{21.9.4}$$

where $\mathcal{Z}$ is the $N\times N$ matrix with matrix elements Z_{nm} and $\boldsymbol{I}$, $\mathbf{s}$, $\mathbf{c}$ are the column vectors with elements I_n, $c_n = \cos kz_n$, and $s_n = \sin k|z_n|$, for $n = -M,\dots,M$. The vector $\boldsymbol{I}$ is *symmetric* about its middle, that is, $I_{-n} = I_n$, and similarly for $\mathbf{c}, \mathbf{s}$. Therefore, we have:

$$\boldsymbol{I} = \begin{bmatrix} I_M \\ \vdots \\ I_1 \\ I_0 \\ I_1 \\ \vdots \\ I_M \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} v_M \\ \vdots \\ v_1 \\ v_0 \\ v_1 \\ \vdots \\ v_M \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_M \\ \vdots \\ c_1 \\ c_0 \\ c_1 \\ \vdots \\ c_M \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_M \\ \vdots \\ s_1 \\ s_0 \\ s_1 \\ \vdots \\ s_M \end{bmatrix} \tag{21.9.5}$$

The matrix $\mathcal{Z}$ is a symmetric Toeplitz matrix because the matrix element Z_{nm} depends only on the difference $|n-m|$. Taking advantage of the Toeplitz nature of $\mathcal{Z}$ and the symmetry of the vectors (21.9.5), the matrix system (21.9.4) can be replaced by one essentially half its size, thus, speeding up the solution. To see this, we partition the vector $\boldsymbol{I}$ into its upper (negative-z), middle, and lower (positive-z) parts:

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{I}_1^R \\ I_0 \\ \boldsymbol{I}_1 \end{bmatrix}, \quad \boldsymbol{I}_1 = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_M \end{bmatrix}, \quad \boldsymbol{I}_1^R = \begin{bmatrix} I_M \\ \vdots \\ I_2 \\ I_1 \end{bmatrix}$$

The upper part $\boldsymbol{I}_1^R$ is the *reverse* of the lower part $\boldsymbol{I}_1$. The reversal operation can be expressed as a matrix operation:

$$\boldsymbol{I}_1^R = J\boldsymbol{I}_1\,, \qquad J = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & ⋰ & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}$$

where J is the $M\times M$ reversing matrix J, that is, the matrix with ones along its antidiagonal. Then, the impedance matrix $\mathcal{Z}$ and Eq. (21.9.4) can be partitioned in a compatible way as follows:

$$\begin{bmatrix} A^R & \mathbf{a}^R & B^R \\ \mathbf{a}^{TR} & a_0 & \mathbf{a}^T \\ B & \mathbf{a} & A \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_1^R \\ I_0 \\ \boldsymbol{I}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^R \\ v_0 \\ \boldsymbol{v}_1 \end{bmatrix} \tag{21.9.6}$$

where we have separated the middle column and row of $\mathcal{Z}$. Because $\mathcal{Z}$ satisfies the reversal invariance condition $\mathcal{Z}(n,m)=\mathcal{Z}(-n,-m)$, the upper-left block A^R will be the reverse of the lower-right block A, and the upper-right, the reverse of the lower-left. Moreover, because $\mathcal{Z}$ is symmetric, we have $A^R = A^T = A$ and $B^R = B^T$.

The reverse of a matrix is obtained by reversing its columns and then reversing its rows, an operation which is equivalent to multiplication by the reversing matrix J from left and right:

$$A^R = JAJ$$

Writing out the three sub-block equations of Eq. (21.9.6), we obtain:

$$\begin{aligned} A^R\boldsymbol{I}_1^R + \mathbf{a}^R I_0 + B^R\boldsymbol{I}_1 &= \boldsymbol{v}_1^R \\ \mathbf{a}^{TR}\boldsymbol{I}_1^R + a_0 I_0 + \mathbf{a}^T\boldsymbol{I}_1 &= v_0 \\ B\boldsymbol{I}_1^R + \mathbf{a}I_0 + A\boldsymbol{I}_1 &= \boldsymbol{v}_1 \end{aligned}$$

But, the first is exactly the reverse of the last, and therefore redundant. Noting that $\mathbf{a}^{TR}\boldsymbol{I}_1^R = \mathbf{a}^T\boldsymbol{I}_1$ and $B\boldsymbol{I}_1^R = BJ\boldsymbol{I}_1$, we obtain the reduced system:

$$\begin{aligned} a_0 I_0 + 2\mathbf{a}^T\boldsymbol{I}_1 &= v_0 \\ \mathbf{a}I_0 + (A + BJ)\boldsymbol{I}_1 &= \boldsymbol{v}_1 \end{aligned}$$

which can be written in the reduced block matrix form:

$$\begin{bmatrix} a_0 & 2\mathbf{a}^T \\ \mathbf{a} & A+BJ \end{bmatrix}\begin{bmatrix} I_0 \\ \boldsymbol{I}_1 \end{bmatrix} = \begin{bmatrix} v_0 \\ \boldsymbol{v}_1 \end{bmatrix} \tag{21.9.7}$$

Thus, we can replace the $N{\times}N$ system (21.9.4) or (21.9.6) by the $(M+1)\times(M+1)$ system (21.9.7) acting only on half-vectors. We will write Eq. (21.9.7) in the following compact form:

$$\boxed{Z\boldsymbol{I} = \boldsymbol{v} = C_1\mathbf{c} + V_0\mathbf{s}} \tag{21.9.8}$$

where Z is constructed from $\mathcal{Z}$ according to (21.9.7) and the vectors are the half-vectors:

$$\boldsymbol{I} = \begin{bmatrix} I_0 \\ I_1 \\ \vdots \\ I_M \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_M \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_M \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_M \end{bmatrix} \tag{21.9.9}$$

Next, we impose the condition that $I_M = 0$ from which the constant C_1 can be determined. This condition can be written vectorially in the form $\mathbf{u}^T\boldsymbol{I} = 0$, where $\mathbf{u}^T = [0,\dots,0,1]$. Solving (21.9.8) for $\boldsymbol{I}$, we obtain:

$$\boxed{\boldsymbol{I} = C_1 Z^{-1}\mathbf{c} + V_0 Z^{-1}\mathbf{s}} \tag{21.9.10}$$

Multiplying both sides by $\mathbf{u}^T$, we obtain the condition:

$$\mathbf{u}^T\boldsymbol{I} = C_1\mathbf{u}^T Z^{-1}\mathbf{c} + V_0\mathbf{u}^T Z^{-1}\mathbf{s} = 0$$

which may be solved for C_1:

$$\boxed{C_1 = -V_0\frac{\mathbf{u}^T Z^{-1}\mathbf{s}}{\mathbf{u}^T Z^{-1}\mathbf{c}}} \tag{21.9.11}$$

The two equations (21.9.10) and (21.9.11) provide the complete solution of the discretized Hallén equation. The MATLAB function `hdelta` implements the above solution procedure. It has usage:

```
[I,z,cnd] = hdelta(L,a,M,ker,basis);     % solve Hallén equation with delta-gap input
```

The string parameter inputs `ker` and `basis` take the possible values:

```
ker = 'e', 'a'                  % exact or approximate kernel
basis = 'p', 't', 'n', 'd'      % pulse, triangular, NEC, or delta-function basis
```

where the choice `basis='d'` applies in this section. This function solves the half system (21.9.8), but returns the full N-dimensional symmetric vector $\boldsymbol{I}$ of Eq. (21.9.5). The quantity $\mathbf{z}$ is the N-dimensional vector of sampled z-points (21.8.2), and `cnd` is the condition number of the matrix Z that is being inverted. The quantities L, a are the antenna length and radius in units of λ, and M has the same meaning as above.

The matrix inversions required in (21.9.10) can be implemented efficiently by MATLAB's backslash operation $\mathcal{Z}\backslash[\mathbf{c},\mathbf{s}]$. The function assumes $V_0 = 1$. Therefore, the input impedance of the antenna will be $Z_0 = V_0/I_0 = 1/I_0$, where I_0 is the middle of the output vector $\boldsymbol{I}$ (i.e., the sample $I(M+1)$ in MATLAB indexing.)

Internally, the function `hdelta` calls the functions `hmat` and `hwrap` to construct the impedance matrices $\mathcal{Z}$ and Z:

```
[Z,B] = hmat(L,a,M,ker,basis);     % construct the (2M + 1)×(2M + 1) Hallén impedance matrix
Zwrap = hwrap(Z);                  % wrap impedance matrix to size (M + 1)×(M + 1)
```

where B is a tridiagonal matrix that is applicable only in the NEC basis case, as explained in Sec. 21.12, and is equal to the identity matrix otherwise.

Fast Toeplitz solvers can also be used, based on the Levinson recursion and fast Cholesky factorizations [48]. However, we found that the built-in linear system solver of MATLAB is much faster for sizes of the order $M = 20$–200.

Example 21.9.1: To clarify the structure of the impedance matrix $\mathcal{Z}$ and show how to wrap it efficiently into the half-size of (21.9.7), consider the case $N = 7$ or $M = 3$. Because $\mathcal{Z}$ is Toeplitz and symmetric, it can be built from the knowledge of its first column or first row. The first column is $\mathcal{Z}_{n,-M} = \mathcal{Z}_{n+M,0}$, for $-M \le n \le M$. Setting $m = n + M$, so that $m = 0, 1, \dots, 2M$, the first column (and first row) consists of the numbers:

$$a_m = \mathcal{Z}_{m,0}\,, \quad m = 0, 1, \dots, 2M \tag{21.9.12}$$

Therefore, the full matrix $\mathcal{Z}$ will have the form:

$$\mathcal{Z} = \left[\begin{array}{ccc|c|ccc} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_1 & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \\ a_2 & a_1 & a_0 & a_1 & a_2 & a_3 & a_4 \\ \hline a_3 & a_2 & a_1 & a_0 & a_1 & a_2 & a_3 \\ \hline a_4 & a_3 & a_2 & a_1 & a_0 & a_1 & a_2 \\ a_5 & a_4 & a_3 & a_2 & a_1 & a_0 & a_1 \\ a_6 & a_5 & a_4 & a_3 & a_2 & a_1 & a_0 \end{array}\right]$$

where we partitioned it as in Eq. (21.9.7), with submatrices:

$$A = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_1 & a_0 & a_1 \\ a_2 & a_1 & a_0 \end{bmatrix}, \quad B = \begin{bmatrix} a_4 & a_3 & a_2 \\ a_5 & a_4 & a_3 \\ a_6 & a_5 & a_4 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Therefore, the wrapped version of $\mathcal{Z}$ will be:

$$\mathcal{Z} = \begin{bmatrix} a_0 & 2\mathbf{a}^T \\ \mathbf{a} & A + BJ \end{bmatrix} = \begin{bmatrix} a_0 & 2a_1 & 2a_2 & 2a_3 \\ a_1 & a_0 + a_2 & a_1 + a_3 & a_2 + a_4 \\ a_2 & a_1 + a_3 & a_0 + a_4 & a_1 + a_5 \\ a_3 & a_2 + a_4 & a_1 + a_5 & a_0 + a_6 \end{bmatrix} \tag{21.9.13}$$

This matrix can be constructed quickly as follows. Once the numbers a_m, $m = 0, 1, \dots, 2M$ are computed, take the first and last $M + 1$ numbers, that is, define the two row vectors:

$$\boldsymbol{f} = [a_0, a_1, a_2, a_3, a_4, a_5, a_6] \quad \Rightarrow \quad \boldsymbol{g} = [a_0, a_1, a_2, a_3]\,, \ \boldsymbol{h} = [a_3, a_4, a_5, a_6]$$

Then, form the Toeplitz matrix whose first column and first row are $\boldsymbol{g}$, and add it to the Hankel matrix whose first column is $\boldsymbol{g}$ and last row is $\boldsymbol{h}$. This is accomplished easily by the built-in MATLAB functions `toeplitz` and `hankel`:

$$\text{toeplitz}\,(\boldsymbol{g}, \boldsymbol{g}) + \text{hankel}\,(\boldsymbol{g}, \boldsymbol{h}) = \begin{bmatrix} 2a_0 & 2a_1 & 2a_2 & 2a_3 \\ 2a_1 & a_0 + a_2 & a_1 + a_3 & a_2 + a_4 \\ 2a_2 & a_1 + a_3 & a_0 + a_4 & a_1 + a_5 \\ 2a_3 & a_2 + a_4 & a_1 + a_5 & a_0 + a_6 \end{bmatrix}$$

Then, replace the first column by half its value. These procedures are incorporated into the function `hwrap`. We note that the full matrix $\mathcal{Z}$ can also be constructed using the function `toeplitz` by:

$$\mathcal{Z} = \text{toeplitz}\,(\boldsymbol{f}, \boldsymbol{f})$$

This is how $\mathcal{Z}$ is constructed by the function `hmat`. □

21.10 Pulse Basis

The delta function discretization scheme described in the previous section yields comparable results to the other bases. Here, we look at the pulse-function basis, which was defined by Eq. (21.8.4) and shown in Fig. 21.8.1. For the *point-matching case*, the matrix elements Z_{nm} are given by Eq. (21.8.9):

$$Z_{nm} = \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n - z')B(z' - z_m)\,dz' = \frac{j\eta}{2\pi}\int_{z_m-\Delta/2}^{z_m+\Delta/2} G(z_n - z')\,dz'$$

or, changing integration variable to $z = z' - z_m$,

$$Z_{nm} = \frac{j\eta}{2\pi}\int_{-\Delta/2}^{\Delta/2} G(z_n - z_m - z)\,dz\,, \quad -M \le n, m \le M \tag{21.10.1}$$

The integration range can be folded in half giving:

$$Z_{nm} = \frac{j\eta}{2\pi}\int_{0}^{\Delta/2} \big[G(z_n - z_m - z) + G(z_n - z_m + z)\big]\,dz \tag{21.10.2}$$

The function `hmat` with the option, `basis=’p’`, may be used to evaluate Z_{nm}, for both the exact and the approximate kernels. A 32-point Gauss-Legendre quadrature integration rule is used to evaluate the integral in (21.10.2). The Hallén matrix equation is again given by (21.9.4):

$$\mathcal{Z}\boldsymbol{I} = \boldsymbol{v} = C_1\mathbf{c} + V_0\mathbf{s} \tag{21.10.3}$$

and may be solved using the techniques of the previous sections. The function `hdelta` solves (21.10.3) with the option, `basis=’p’`.

For the *Galerkin case*, the matrix equation (21.10.3) has the same form but with some redefinitions of the quantities $\mathcal{Z}$, $\mathbf{c}$, $\mathbf{s}$. The matrix elements Z_{nm} are given by (21.8.10):

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h}\int_{-h}^{h} B(z - z_n)G(z - z')B(z' - z_m)\,dz\,dz' = \\ &= \frac{j\eta}{2\pi}\int_{z_n-\Delta/2}^{z_n+\Delta/2}\int_{z_m-\Delta/2}^{z_m+\Delta/2} G(z - z')\,dz\,dz' \\ &= \frac{j\eta}{2\pi}\int_{-\Delta/2}^{\Delta/2}\int_{-\Delta/2}^{\Delta/2} G(z_n - z_m + x - x')\,dx\,dx' \end{aligned}$$

where we changed variables to $x = z - z_n$ and $x' = z' - z_m$. Defining a new integration variable $z = x - x'$, then one of the remaining integrations can be done giving:

$$Z_{nm} = \frac{j\eta}{2\pi}\int_{-\Delta}^{\Delta} (\Delta - |z|)G(z_n - z_m + z)\,dz\,, \quad -M \le n, m \le M \tag{21.10.4}$$

Similarly, the right-hand side vector $\boldsymbol{v}$ of (21.10.3) has components:

$$\begin{aligned} v_n &= \int_{-h}^{h} B(z - z_n)V(z)\,dz = \int_{-\Delta/2}^{\Delta/2} V(z_n + x)\,dx \\ &= C_1\int_{-\Delta/2}^{\Delta/2} \cos k(z_n + x)\,dx + V_0\int_{-\Delta/2}^{\Delta/2} \sin k|z_n + x|\,dx \equiv C_1 c_n + V_0 s_n \end{aligned}$$

where the individual terms c_n, s_n can be calculated explicitly:

$$\begin{aligned} c_n &= \int_{-\Delta/2}^{\Delta/2} \cos k(z_n + x)\,dx = \frac{2}{k}\sin\Big(\frac{k\Delta}{2}\Big)\cos kz_n \\ s_n &= \int_{-\Delta/2}^{\Delta/2} \sin k|z_n + x|\,dx = \frac{4}{k}\sin^2\Big(\frac{k\Delta}{4}\Big)\delta(n) + \frac{2}{k}\sin\Big(\frac{k\Delta}{2}\Big)\sin k|z_n| \end{aligned} \tag{21.10.5}$$

where $\delta(n)$ is the Kronecker delta, and $z_n = n\Delta$, for $-M \le n \le M$. The Galerkin method yields very comparable results to the point matching case.

Some examples of the point-matching case with pulse-function basis were shown in Figs. 21.6.1 and 21.6.2. Some further examples are given below based on [1253,1254].

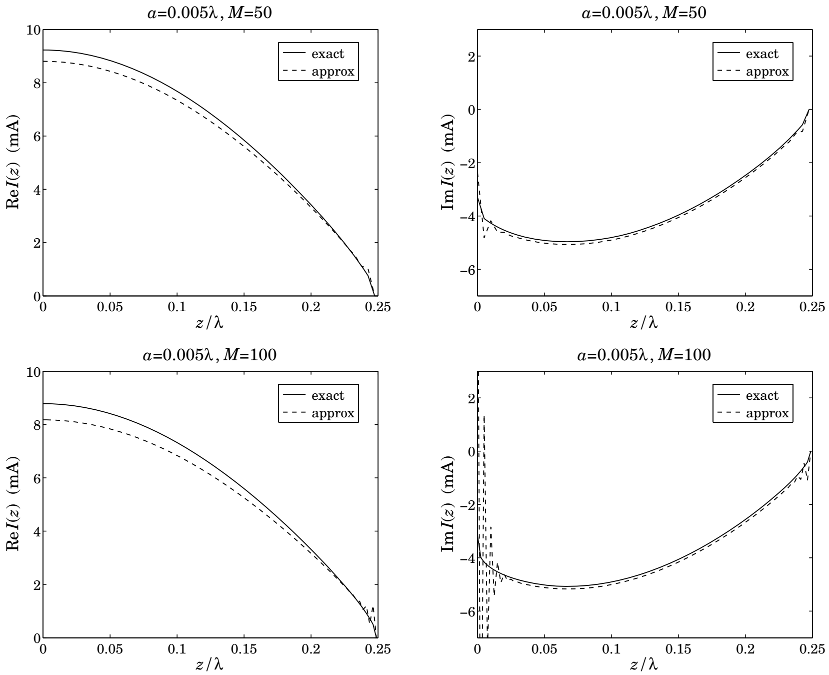


Fig. 21.10.1 Real and imaginary parts of $I(z)$ for half-wave dipole, $M = 50$ and $M = 100$.

Fig. 21.10.1 compares the solutions of (21.10.3) using the exact and the approximate kernels for a dipole of length $l = 0.5\lambda$ and radius $a = 0.005\lambda$.

The upper two graphs show the real and imaginary parts of the current samples I_n (joined here by straight lines) using $M = 50$, which corresponds to $N = 2M + 1 = 101$ segments. The lower two graphs show the case of $M = 100$. We note that for the approximate kernel, the solution oscillates wildly near the center and the end-points of the antenna—a behavior attributed to the non-existence of solutions of the Hallén equation Eq. (21.8.1) in this case. Fig. 21.10.2 depicts the case of a half-wave dipole with a larger radius $a = 0.008\lambda$ and $M = 100$, for which the oscillations get worse.

We have not superimposed King's three-term fit because it is virtually indistinguishable from the exact-kernel solution. The graphs, including the 3-term fit, may be generated by the following example MATLAB code:

```
L=0.5; a=0.005; M=100;

[Ie,z] = hdelta(L,a,M,'e','p');      % solution of Hallén equation with exact kernel
Ia = hdelta(L,a,M,'a','p');          % solution of Hallén equation with approximate kernel
A = kingfit(L,Ie,z,3);               % three-term coefficients fitted to exact-kernel solution
I3 = kingeval(L,A,z);                % evaluate 3-term current at the sample points z

figure; plot(z,real(Ie),'-', z,real(Ia),':', z,real(I3),'--');
figure; plot(z,imag(Ie),'-', z,imag(Ia),':', z,imag(I3),'--');
```

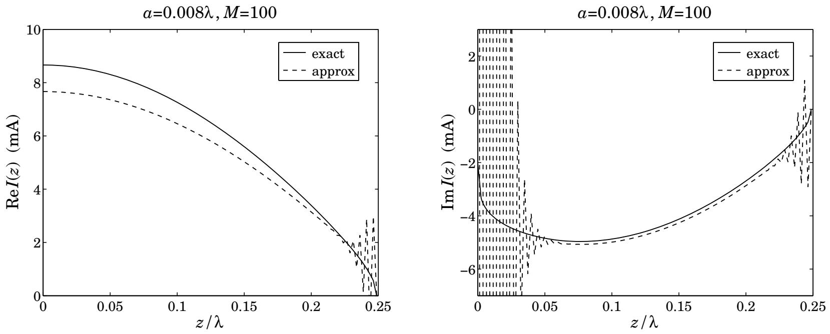


Fig. 21.10.2 Current on half-wave dipole of radius $a = 0.008\lambda$ and $M = 100$.

In Fig. 21.10.3, we compare how well the logarithmic behavior of Eq. (21.8.11) near $z = 0$ fits the computed current $I(z_n)$.

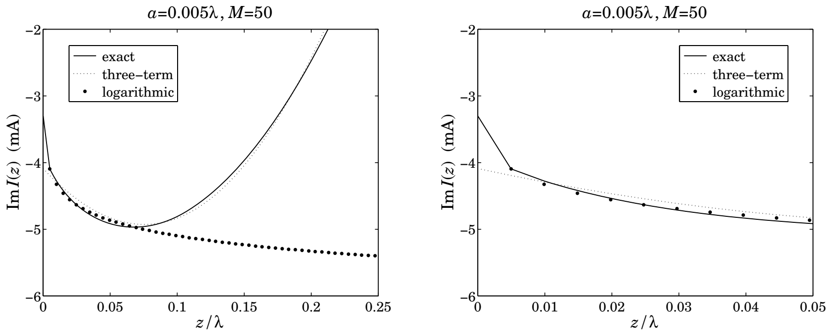


Fig. 21.10.3 Logarithmic behavior of the solution near $z = 0$.

To make the comparison, we adjust the asymptotic form (21.8.11) to match the computed value at the the closest z-point to zero, that is, at $z_1 = \Delta$. The adjusted asymptotic current is then,

$$I_{\log}(z) \simeq -j\,\frac{4kaV_0}{\eta}\,\ln\left|\frac{z}{z_1}\right| + I(z_1)\,,\quad z \simeq 0 \tag{21.10.6}$$

so that it satisfies $I_{\log}(z_1) = I(z_1)$. For the case $a = 0.005\lambda$ and $M = 50$, Fig. 21.10.3 plots the imaginary part of the computed current $I(z)$ based on the exact kernel, together with the corresponding three-term fit, and the asymptotic current $I_{\log}(z)$.

The graph on the left is over the interval $0 \le z \le 0.25\lambda$, whereas the graph on the right shows only the narrow interval $0 \le z \le 0.05\lambda$, and both graphs use an expanded vertical scale compared to that of Fig. 21.10.1. The logarithmic behavior is evident. We note also that the three-term fit agrees very well with the computed current except in a narrow interval about $z = 0$.

Finally, we look at the stability of the numerical solutions based on the exact and approximate kernels as the number of segments M increases. Fig. 21.10.4 depicts the calculated input admittance $Y_0 = I_0/V_0$ as a function of the number of segments M.

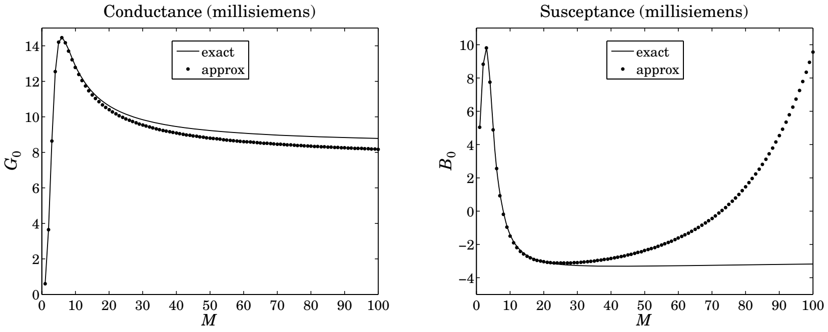


Fig. 21.10.4 Input admittance of half-wave dipole vs. number of segments M.

The conductance $G_0 = \mathrm{Re}(Y_0)$ converges for both the exact and approximate kernel solutions. However, the susceptance $B_0 = \mathrm{Im}(Y_0)$ diverges for the approximate kernel case, and converges for the exact kernel—the two agreeing well only for $M \lesssim 30$.

Fig. 21.10.5 depicts the dependence of the condition number of the Hallén impedance matrix $\mathcal{Z}$ on the number of segments M. We note the relative well-conditioning of the exact-kernel case. The above graphs may be computed with the following MATLAB code:

```
L=0.5; a=0.005;

for M=1:1:100,
   [I,z,cnd] = hdelta(L,a,M,’e’,’p’); Ye(M) = I(M+1); ce(M) = cnd;
   [I,z,cnd] = hdelta(L,a,M,’a’,’p’); Ya(M) = I(M+1); ca(M) = cnd;
end

M=1:1:100;

figure; plot(M,real(Ye),’-’, M,real(Ya),’.’);
figure; plot(M,imag(Ye),’-’, M,imag(Ya),’.’);
figure; semilogy(M,ce,’-’, M,ca,’--’);
```

For very small antenna radius, the exact and approximate kernels produce essentially the same current solutions. Fig. 21.10.6 shows the input impedance $Z_0 = R + jX = V_0/I_0$ versus the antenna length in the interval $0.3\lambda \le l \le 0.7\lambda$, and for the two radii $a = 0.0005\lambda$ and 0.000001λ.

The calculation was carried out using a pulse basis and the approximate kernel with $M = 100$ segments, but one could just as well have used the exact kernel or a triangular basis with very little difference in the results. The graphs are similar to those of Fig. 16.3.1 and show the resonant length at about 0.48λ–0.49λ. The graphs were produced with the MATLAB code:

```
L = linspace(0.3,0.7,41);
a = [0.0005, 0.00001];
```

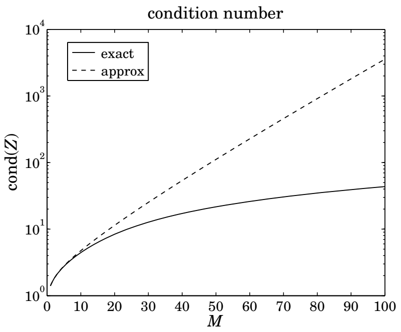


Fig. 21.10.5 Condition number of Hallén impedance matrix vs. number of segments M.

```
M = 100;

for i=1:length(L),
   for k=1:length(a),
      I = hdelta(L(i),a(k),M,’a’,’p’); Z(i,k) = 1/I(M+1);
   end
end

figure; plot(L, real(Z(:,1)), ’-’, L, real(Z(:,2)), ’--’);
figure; plot(L, imag(Z(:,1)), ’-’, L, imag(Z(:,2)), ’--’);
```

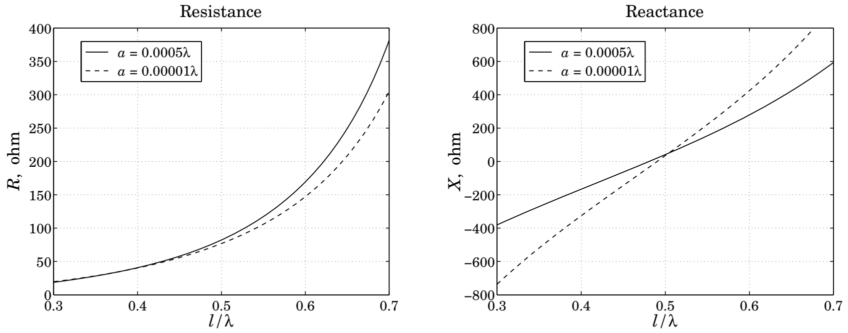


Fig. 21.10.6 Input impedance versus antenna length.

21.11 Triangular Basis

For the triangular basis, the current expansion (21.8.3), reads

$$I(z') = \sum_{m=-M}^{M} I_m B(z' - z_m) \tag{21.11.1}$$

where $B(z)$ is the triangular function:

$$B(z)=\begin{cases}1-\dfrac{|z|}{\Delta}\,, & \text{if } |z|\le\Delta\\[4pt] 0\,, & \text{otherwise}\end{cases} \tag{21.11.2}$$

The triangular basis functions, depicted in Fig. 21.11.1, have duration 2Δ, that is, twice as long as the pulse case. The linear combination of (21.11.1) is equivalent to connecting the sample values I_m by straight line segments.

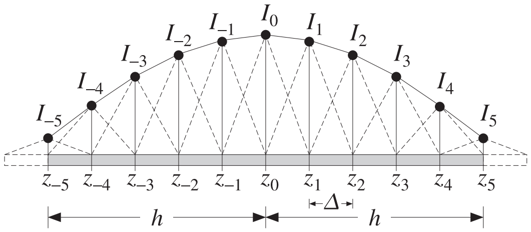


Fig. 21.11.1 Triangular basis functions with $N = 11$, $M = 5$.

The antenna is divided into $N = 2M$ segments of width $\Delta = l/2M = h/M$, and the $(2M+1)$ points $z_m = m\Delta$, $-M \le m \le M$, lie at the end points of these segments. As depicted in Fig. 21.11.1, the influence of the last current sample I_M at z_M (and I_{-M} at z_{-M}) may be thought of as extending beyond the end of the antenna by an additional segment. The vanishing of the current at the ends of the antenna is enforced by the conditions $I_M = I_{-M} = 0$.

For *point-matching*, the Hallén impedance matrix (21.8.9) will be given as follows:

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n-z')B(z'-z_m)\,dz' \\ &= \frac{j\eta}{2\pi}\int_{z_m-\Delta}^{z_m+\Delta}\left(1-\frac{|z'-z_m|}{\Delta}\right)G(z_n-z')\,dz' \\ &= \frac{j\eta}{2\pi}\int_{-\Delta}^{\Delta}\left(1-\frac{|z|}{\Delta}\right)G(z_n-z_m-z)\,dz \end{aligned} \tag{21.11.3}$$

which differs by a factor of Δ from the Galerkin case of the pulse basis, Eq. (21.10.4). The Hallén matrix equation remains the same as (21.9.4):

$$Z\boldsymbol{I} = \boldsymbol{v} = C_1\mathbf{c} + V_0\mathbf{s} \tag{21.11.4}$$

with the same right-hand vectors, that is, $c_n = \cos kz_n$ and $s_n = \sin k|z_n|$, $-M \le n \le M$. It may be solved by the function `hmat`, called with the option, `basis='t'`. A 32-point Gauss-Legendre quadrature is used to compute the integrals in (21.11.3).

The exact and approximate kernel current solutions and input admittance exhibit the same behavior in the triangular basis as in the pulse basis.

Using the exact kernel, the triangular basis converges faster than the pulse basis as the number of segments M increases. Fig. 21.11.2 compares the convergence of the input admittance of a half-wave dipole with radius $a = 0.005\lambda$ for the two bases.

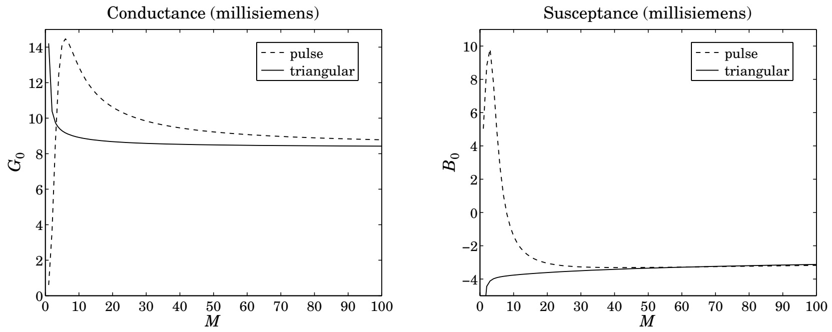


Fig. 21.11.2 Input admittance computed with pulse and triangular basis functions.

21.12 NEC Sinusoidal Basis

The Numerical Electromagnetics Code (NEC) is a widely used public-domain program for modeling antennas and other structures [1227]. The program solves Pocklington's equation using point-matching and a spline-like sinusoidal basis. A similar basis was originally used by [1211]. For a linear antenna, the basis is defined by:

$$B(z)=\begin{cases} A_- + B_-\sin k(z+\Delta) + C_-\cos k(z+\Delta)\,, & -\dfrac{3\Delta}{2}\le z\le -\dfrac{\Delta}{2} \\[6pt] A_0 + B_0\sin kz + C_0\cos kz\,, & -\dfrac{\Delta}{2}\le z\le \dfrac{\Delta}{2} \\[6pt] A_+ + B_+\sin k(z-\Delta) + C_+\cos k(z-\Delta)\,, & \dfrac{\Delta}{2}\le z\le \dfrac{3\Delta}{2} \end{cases} \tag{21.12.1}$$

where Δ is the segment length defined as usual by $\Delta = 2h/(2M+1)$. The basis function $B(z)$ is shown on the left of Fig. 21.12.1. It extends over three consecutive segments.

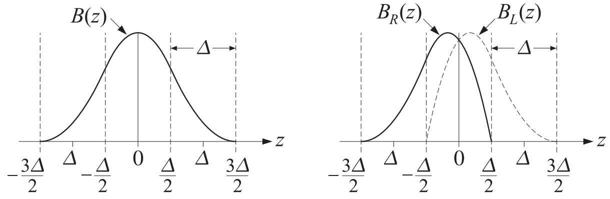


Fig. 21.12.1 NEC sinusoidal basis functions.

The A, B, C coefficients are determined by imposing the spline-like conditions that (a) the three pieces of $B(z)$ join continuously, as do their slopes, at the points $z = \pm\Delta/2$, (b) at the end-points $z = \pm 3\Delta/2$, the function $B(z)$ and its derivative vanish, and (c) at the center $z = 0$, $B(0)$ is normalized to one. These conditions provide nine equations for the nine coefficients, with solution:

$$A_0 = \frac{1 - 2c_\Delta^2}{1 + c_\Delta - 2c_\Delta^2}\,, \quad B_0 = 0\,, \quad C_0 = \frac{c_\Delta}{1 + c_\Delta - 2c_\Delta^2} \tag{21.12.2}$$

$$A_\pm = \frac{1}{2(1 + c_\Delta - 2c_\Delta^2)}\,,\; B_\pm = \mp\frac{s_\Delta}{2(1 + c_\Delta - 2c_\Delta^2)}\,,\; C_\pm = -\frac{c_\Delta}{2(1 + c_\Delta - 2c_\Delta^2)} \tag{21.12.3}$$

where c_Δ, s_Δ are shorthands for the quantities:

$$c_\Delta = \cos\left(\frac{k\Delta}{2}\right)\,, \quad s_\Delta = \sin\left(\frac{k\Delta}{2}\right) \tag{21.12.4}$$

The current expansion in terms of this basis is as follows:

$$I(z) = \sum_{m=-M}^{M} b_m B(z - z_m) \tag{21.12.5}$$

where $z_m = m\Delta$, $-M \le m \le M$ and the coefficients b_m are to be determined.

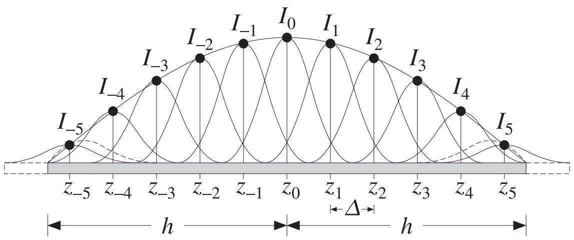


Fig. 21.12.2 NEC sinusoidal basis with $N = 11$, $M = 5$.

Fig. 21.12.2 depicts (21.12.5). Actually, the NEC basis handles the two end segments centered at $z_{\pm M}$ separately [1227] and uses the following expansion instead of (21.12.5):

$$I(z) = b_{-M}B_L(z - z_{-M}) + \sum_{m=-(M-1)}^{(M-1)} b_m B(z - z_m) + b_M B_R(z - z_M) \tag{21.12.6}$$

where $B_R(z)$ and $B_L(z)$ are given by:

$$B_R(z) = \begin{cases} A_-^R + B_-^R \sin k(z+\Delta) + C_-^R \cos k(z+\Delta)\,, & -\dfrac{3\Delta}{2} \le z \le -\dfrac{\Delta}{2} \\[2ex] A_0^R + B_0^R \sin kz + C_0^R \cos kz\,, & -\dfrac{\Delta}{2} \le z \le \dfrac{\Delta}{2} \end{cases} \tag{21.12.7}$$

and $B_L(z) = B_R(-z)$, where the coefficients are determined by requiring (a) the continuity of the two pieces and their derivatives at $z = -\Delta/2$, (b) the vanishing of $B_R(z)$ at $z = \Delta/2$, (c) the vanishing of $B_R(z)$ and its derivative at $z = -3\Delta/2$, and (d) the normalization condition $B_R(0) = 1$. These give:

$$\begin{aligned} A_-^R &= \frac{1}{1 + 3c_\Delta - 4c_\Delta^2}\,, \quad B_-^R = \frac{s_\Delta}{1 + 3c_\Delta - 4c_\Delta^2}\,, \quad C_-^R = -\frac{c_\Delta}{1 + 3c_\Delta - 4c_\Delta^2} \\ A_0^R &= \frac{1 - 4c_\Delta^2}{1 + 3c_\Delta - 4c_\Delta^2}\,, \quad B_0^R = -\frac{s_\Delta}{1 + 3c_\Delta - 4c_\Delta^2}\,, \quad C_0^R = \frac{3c_\Delta}{1 + 3c_\Delta - 4c_\Delta^2} \end{aligned} \tag{21.12.8}$$

The functions $B_L(z)$ and $B_R(z)$ are depicted on the right of Fig. 21.12.1, and are shown as dashed curves at the antenna ends in Fig. 21.12.2. Because of the end conditions $I_{\pm M} = 0$, it makes very little difference whether one uses (21.12.5) or (21.12.6). We will use the former.

Because $B(z)$ straddles three adjacent segments and $B(\pm\Delta) \ne 0$, the value of the current at a sample point z_n will receive contributions from the two adjacent samples z_{n-1} and z_{n+1}. For $-M < n < M$, we have from (21.12.5):

$$\begin{aligned} I_n = I(z_n) &= b_{n-1}B(z_n - z_{n-1}) + b_n B(z_n - z_n) + b_{n+1}B(z_n - z_{n+1}) \\ &= b_{n-1}B(\Delta) + b_n B(0) + b_{n+1}B(-\Delta) \end{aligned}$$

or, noting that $B(0) = 1$ and defining $\beta = B(\Delta) = B(-\Delta)$,

$$I_n = \beta b_{n-1} + b_n + \beta b_{n+1}\,, \quad -M < n < M \tag{21.12.9}$$

and for the end points:

$$I_{-M} = b_{-M} + \beta b_{-M+1}\,, \quad I_M = \beta b_{M-1} + b_M \tag{21.12.10}$$

Eqs. (21.12.9) and (21.12.10) can be arranged into a tridiagonal matrix, for example,

$$\boldsymbol{I} = \begin{bmatrix} I_{-3} \\ I_{-2} \\ I_{-1} \\ I_0 \\ I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1 & \beta & 0 & 0 & 0 & 0 & 0 \\ \beta & 1 & \beta & 0 & 0 & 0 & 0 \\ 0 & \beta & 1 & \beta & 0 & 0 & 0 \\ 0 & 0 & \beta & 1 & \beta & 0 & 0 \\ 0 & 0 & 0 & \beta & 1 & \beta & 0 \\ 0 & 0 & 0 & 0 & \beta & 1 & \beta \\ 0 & 0 & 0 & 0 & 0 & \beta & 1 \end{bmatrix} \begin{bmatrix} b_{-3} \\ b_{-2} \\ b_{-1} \\ b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \equiv \mathcal{B}\boldsymbol{b} \tag{21.12.11}$$

It follows from Eq. (21.12.1) that the parameter β is equal to $A_+ + C_+$, or,

$$\beta = B(\Delta) = \frac{1 - c_\Delta}{2(1 + c_\Delta - 2c_\Delta^2)} \tag{21.12.12}$$

For the pulse and triangular bases, the matrix $\mathcal{B}$ is equal to the identity matrix because in these cases $\beta = B(\Delta) = 0$.

Next, we determine the solution of Hallén's equation with point matching. Inserting the expansion (21.12.5) into (21.8.1), and evaluating it at the sample point z_n, we obtain the matrix equation for the coefficients b_m:

$$\sum_{m=-M}^{M} Z_{nm} b_m = C_1 \cos kz_n + V_0 \sin k|z_n|\,, \quad -M \le n \le M \tag{21.12.13}$$

or, written compactly,

$$\mathcal{Z}\boldsymbol{b} = C_1\mathbf{c} + V_0\mathbf{s} \tag{21.12.14}$$

where the matrix elements Z_{nm} are given by:

$$\begin{aligned} Z_{nm} &= \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n - z')B(z' - z_m)\,dz' \\ &= \frac{j\eta}{2\pi}\int_{-3\Delta/2}^{3\Delta/2} B(z)G(z_n - z_m - z)\,dy \end{aligned} \tag{21.12.15}$$

Substituting (21.12.11) into (21.12.14), we obtain,

$$\mathcal{Z}\mathcal{B}^{-1}\boldsymbol{I} = C_1\mathbf{c} + V_0\mathbf{s} \tag{21.12.16}$$

This has the same form as (21.9.4) with the replacement $\mathcal{Z} \to \mathcal{Z}\mathcal{B}^{-1}$ and therefore, it can be solved by the same method. The MATLAB function `hdelta` implements the solution with the option, `basis='n'`.

The function `hmat` calculates $\mathcal{Z}$, and it also optionally, outputs the matrix $\mathcal{B}$, which is required in (21.12.16). The integral in (21.12.15) is calculated with Gauss-Legendre quadrature. The basis function $B(z)$ may be evaluated with the help of the MATLAB function `hbasis`, with usage:

```
B = hbasis(z,Delta,basis);    % evaluate basis function B(z) at a vector of z's
```

where `Delta` is the value of Δ and the string `basis` can take the possible values `'p'`, `'t'`, `'n'`, `'nR'`, `'nL'`, for pulse, triangular, and NEC basis (the last two options evaluate the rightmost and leftmost NEC basis functions.)

The NEC basis for the Hallén equation has very comparable performance to the pulse basis with point matching—the two agreeing to within one percent or so, and hence, we do not give any further examples.

21.13 Hallén's Equation for Arbitrary Incident Field

Here, we discuss the solution of Hallén's equation (21.3.11) with arbitrary incident field:

$$\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = C_1 e^{jkz} + C_2 e^{-jkz} + \int_{-h}^{h} F(z-z')E_{\text{in}}(z')\,dz' \tag{21.13.1}$$

In applying the method of moments, we expand the current and the incident field with respect to a given basis $B(z)$:

$$\begin{aligned} I(z') &= \sum_{m=-M}^{M} I_m B(z'-z_m) \\ E_{\text{in}}(z') &= \sum_{m=-M}^{M} E_m B(z'-z_m) \end{aligned} \tag{21.13.2}$$

We consider only the point-matching case. Sampled at the points $z_n = n\Delta z$, the convolution of the incident field with the Green's function $F(z)$ becomes:

$$\int_{-h}^{h} F(z_n - z')E_{\text{in}}(z')\,dz' = \sum_{m=-M}^{M} E_m \int_{-h}^{h} F(z_n - z')B(z'-z_m)\,dz'$$

We define the Green's matrix F_{nm}, for $-M \le n, m \le M$ by

$$F_{nm} = \int_{-h}^{h} F(z_n - z')B(z'-z_m)\,dz' = \int_{-h}^{h} F(z_n - z_m - z)B(z)\,dz \tag{21.13.3}$$

It follows that the discretized Hallén equation (21.13.1) takes the form:

$$\sum_{m=-M}^{M} Z_{nm} I_m = C_1 e^{jkz_n} + C_2 e^{-jkz_n} + \sum_{m=-M}^{M} F_{nm} E_m \tag{21.13.4}$$

where Z_{nm} are defined as in the previous sections:

$$Z_{nm} = \int_{-h}^{h} G(z_n - z')B(z'-z_m)\,dz' = \int_{-h}^{h} G(z_n - z_m - z)B(z)\,dz$$

Eq. (21.13.4) can be written in the compact form:

$$\mathcal{Z}\boldsymbol{I} = C_1\mathbf{s}_1 + C_2\mathbf{s}_2 + F\boldsymbol{E} \tag{21.13.5}$$

where $\mathbf{s}_1$ and $\mathbf{s}_2$ are column vectors with elements $s_1(n) = e^{jkz_n}$ and $s_2(n) = e^{-jkz_n}$. Defining the $N\times 2$ matrix $S = [\mathbf{s}_1, \mathbf{s}_2]$ and the two-dimensional column vector of constants $\boldsymbol{C} = [C_1, C_2]^T$, we write Eq. (21.13.5) in the form:

$$\mathcal{Z}\boldsymbol{I} = S\boldsymbol{C} + F\boldsymbol{E} \tag{21.13.6}$$

For the NEC basis, the expansion (21.13.2) has the modified form:

$$I(z') = \sum_{m=-M}^{M} b_m B(z'-z_m)\,, \quad E_{\text{in}}(z') = \sum_{m=-M}^{M} e_m B(z'-z_m) \tag{21.13.7}$$

where the vectors of the coefficients b_m and e_m are related to the values of the current and field, I_m and E_m, at the sample points z_m via the tridiagonal matrix $\mathcal{B}$ of (21.12.11):

$$\boldsymbol{I} = \mathcal{B}\boldsymbol{b}, \quad \boldsymbol{E} = \mathcal{B}\boldsymbol{e} \tag{21.13.8}$$

The discretized Hallén equation now takes the form:

$$\mathcal{Z}\boldsymbol{b} = S\boldsymbol{C} + F\boldsymbol{e} \tag{21.13.9}$$

and expressed in terms of $\boldsymbol{I}$ and $\boldsymbol{E}$:

$$\mathcal{Z}\mathcal{B}^{-1}\boldsymbol{I} = S\boldsymbol{C} + F\mathcal{B}^{-1}\boldsymbol{E} \tag{21.13.10}$$

For the pulse, triangular, and delta bases $\mathcal{B}$ is replaced by the identity matrix. It is not possible to wrap this equation in half because $\boldsymbol{E}$ is not necessarily symmetric about its middle. The constants $\boldsymbol{C}$ must be found by imposing the two independent end conditions $I(z_M) = I(-z_M) = 0$. These conditions can be written compactly as:

$$U^T\boldsymbol{I} = 0$$

where $U = [\mathbf{u}_{\text{top}}, \mathbf{u}_{\text{bot}}]$ and $\mathbf{u}_{\text{top}} = [1, 0, \dots, 0]^T$ selects the top entry of the vector $\boldsymbol{I}$, while $\mathbf{u}_{\text{bot}} = [0, \dots, 0, 1]^T$ selects the bottom entry. Solving for $\boldsymbol{I}$, we have:

$$\boxed{\boldsymbol{I} = \mathcal{B}\mathcal{Z}^{-1}S\boldsymbol{C} + \mathcal{B}\mathcal{Z}^{-1}F\mathcal{B}^{-1}\boldsymbol{E}} \tag{21.13.11}$$

Multiplying from the left by the matrix U^T, we obtain the condition:

$$U^T\boldsymbol{I} = U^T\mathcal{B}\mathcal{Z}^{-1}S\boldsymbol{C} + U^T\mathcal{B}\mathcal{Z}^{-1}F\mathcal{B}^{-1}\boldsymbol{E} = 0$$

which may be solved for $\boldsymbol{C}$:

$$\boxed{\boldsymbol{C} = -(U^T\mathcal{B}\mathcal{Z}^{-1}S)^{-1}(U^T\mathcal{B}\mathcal{Z}^{-1}F\mathcal{B}^{-1}\boldsymbol{E})} \tag{21.13.12}$$

Eqs. (21.13.11) and (21.13.12) describe the complete solution of the discrete Hallén equation (21.13.10). The MATLAB function `hfield` implements the solution, with usage:

```
[I,z,cnd] = hfield(L,a,E,ker,basis);   % Hallen's equation with arbitrary incident E-field
```

where instead of the parameter M, it has as input the vector $\boldsymbol{E}$ of the samples of the incident field. The dimension $N = 2M+1$ is extracted from the length of $\boldsymbol{E}$. The strings `ker` and `basis` have the same meaning as for the function `hdelta`.

The functions `hdelta` and `hfield` produce practically identical output in the delta-gap case, that is, when the incident field is:

$$\boldsymbol{E} = [\underbrace{0,\,0,\,\dots,\,0}_{M\text{ zeros}},\ \frac{1}{\Delta},\ \underbrace{0,\,\dots,0,0}_{M\text{ zeros}}]^T \tag{21.13.13}$$

The middle entry imitates the delta-gap $V_0\delta(z)\simeq V_0/\Delta$ near $z=0$. For the case of a field incident at a polar angle θ as in Eq. (21.2.4), the sampled vector $\boldsymbol{E}$ will have entries:

$$E_n = E_0\sin\theta\, e^{jkz_n\cos\theta}\,,\quad z_n = n\Delta\,,\quad -M\le n\le M \tag{21.13.14}$$

Fig. 21.13.1 compares the current solutions using `hdelta` and `hfield` with the input of Eq. (21.13.13), for the case of a half-wave dipole of radius $a = 0.005\lambda$, using a pulse basis with $M = 50$ and the exact kernel.

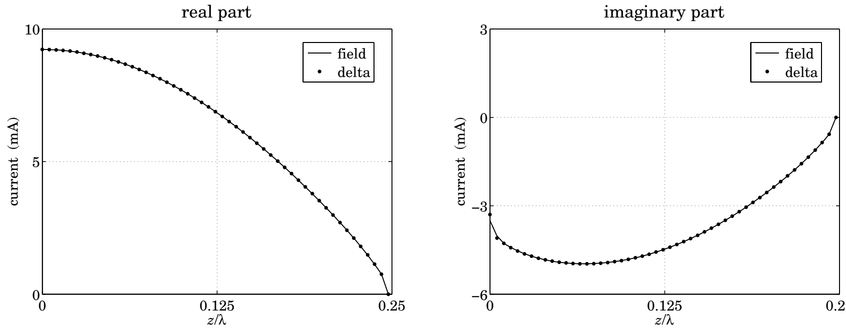


Fig. 21.13.1 Comparison of delta-gap and field solutions for a half-wave dipole.

For larger values of M, the results of the two methods are indistinguishable. The following example MATLAB code can be used to generate these graphs:

```
L=0.5; a=0.005; M = 30; D = L/(2*M+1);
E = zeros(2*M+1,1); E(M+1)=1/D;               % define E by Eq. (21.13.13)

[If,z] = hfield(L,a,E,'e','p');               % field input
[Id,z] = hdelta(L,a,M,'e','p');               % delta-gap input

figure; plot(z,real(If),'-', z,real(Id),'.');
figure; plot(z,imag(If),'-', z,imag(Id),'.');
```

Next, we discuss the computation of the matrix elements F_{nm} for different basis functions. It follows from Eq. (21.13.3) and the even-ness of $B(z)$ that F_{nm} is a Toeplitz and symmetric matrix and, therefore, it depends on n,m through the difference $|n-m|$. Thus, it can be constructed by $F_{nm} = f_{|n-m|}$, where f_m is given by

$$f_m = \int_{-h}^{h} F(m\Delta - z)B(z)\,dz\,,\quad m = 0,1,\dots,2M \tag{21.13.15}$$

Once the vector $\boldsymbol{f} = [f_0, f_1, \dots, f_{2M}]$ is computed, the matrix F can be constructed as a Toeplitz matrix whose first row or first column is $\boldsymbol{f}$, by calling the function `toeplitz`:

$$F = \text{toeplitz}(\boldsymbol{f},\boldsymbol{f})$$

For the Green's function $F(z)$ we choose,

$$F(z) = \sin k|z|$$

although any of those listed in Eq. (21.3.8) could have been chosen. The integrals in (21.13.15) can be done in closed form resulting in the following expressions for f_m in the various bases. For the *delta-function* basis $B(z) = \delta(z)\Delta$, we have:

$$f_m = F(m\Delta)\Delta = \sin(km\Delta)\Delta\,,\quad 0\le m\le 2M \tag{21.13.16}$$

For the *pulse* basis, f_m is given by

$$f_m = \int_{-\Delta/2}^{\Delta/2}\sin\bigl(k|m\Delta - z|\bigr)dz \tag{21.13.17}$$

which gives:

$$\begin{aligned} f_m &= \frac{2}{k}\sin\Bigl(\frac{k\Delta}{2}\Bigr)\sin(km\Delta)\,,\quad 1\le m\le 2M \\ f_0 &= \frac{2}{k}\Bigl(1-\cos\frac{k\Delta}{2}\Bigr) \end{aligned} \tag{21.13.18}$$

For the *triangular* basis, we have:

$$f_m = \int_{-\Delta}^{\Delta}\Bigl(1 - \frac{|z|}{\Delta}\Bigr)\sin\bigl(k|m\Delta - z|\bigr)dz$$

with the result:

$$\begin{aligned} f_m &= \frac{2(1-\cos k\Delta)}{k^2\Delta}\sin(km\Delta)\,,\quad 1\le m\le 2M \\ f_0 &= \frac{2(k\Delta - \sin k\Delta)}{k^2\Delta} \end{aligned} \tag{21.13.19}$$

For the *NEC basis*, f_m is defined by

$$f_m = \int_{-3\Delta/2}^{3\Delta/2} B(z)\sin\big(k|m\Delta - z|\big)\,dz$$

where $B(z)$ is given by (21.12.1), and we obtain:

$$\begin{aligned}
f_m &= \frac{\Delta}{2D}\left(\cos\frac{k\Delta}{2} - \cos\frac{3k\Delta}{2}\right)\sin(km\Delta)\,, \quad 2 \le m \le 2M \\
f_0 &= \frac{2}{kD}\left[\cos\frac{k\Delta}{2} - \cos k\Delta + \frac{k\Delta}{4}\sin\frac{3k\Delta}{2}\right] \\
f_1 &= \frac{1}{kD}\left[1 - \cos\frac{k\Delta}{2} + \frac{k\Delta}{4}\left(\sin\frac{k\Delta}{2} + \sin\frac{3k\Delta}{2} - \sin\frac{5k\Delta}{2}\right)\right]
\end{aligned} \tag{21.13.20}$$

where D is the normalization factor:

$$D = 1 + \cos\frac{k\Delta}{2} - 2\cos^2\frac{k\Delta}{2} = \cos\frac{k\Delta}{2} - \cos k\Delta$$

As in the delta-gap case, the pulse and NEC bases give almost identical results, while the triangular basis converges the fastest. For large M, all bases produce virtually the same result if the exact kernel is used. The approximate kernel solutions suffer from the same type of oscillations as in the delta-gap case.

Fig. 21.13.2 shows the current induced on a half-wave dipole antenna of radius $a = 0.005\lambda$ by a plane wave incident from broadside. Setting $\theta = 90^{\rm o}$ in (21.13.14) results into constant E-field samples $E_m = E_0$, for $-M \le m \le M$. To illustrate the convergence properties, the current was computed using a triangular basis with $M = 30$, and then using a pulse basis with successively increasing values of $M = 30,\ 50,\ 100$, converging to the triangular case. The $M = 50, 100$ cases for a triangular basis are not shown because they are indistinguishable from the $M = 30$ case.

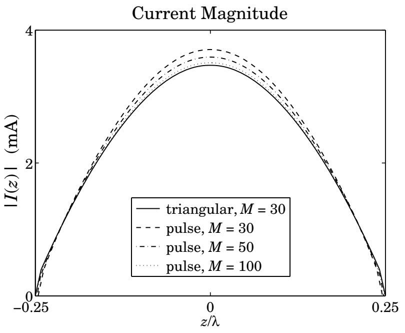


Fig. 21.13.2 Current on half-wave dipole induced by plane wave incident from broadside.

21.14 Solving Pocklington's Equation

In this section, we discuss the numerical solution of Pocklington' s equation (21.1.9) for arbitrary incident field. Rearranging some constants, we can write it in the form:

$$(\partial_z^2 + k^2)\,\frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' = 2kE_{\rm in}(z) \tag{21.14.1}$$

where $G(z)$ is the exact or the approximate kernel. Eq. (21.14.1) must be solved subject to the end-point conditions $I(\pm h) = 0$. Its numerical solution has been studied extensively [1207–1263].

In our method of solution we use point matching and, following Ref. [1226], apply a finite-difference approximation to the Helmholtz operator $(\partial_z^2 + k^2)$ to convert (21.14.1) into a matrix equation for the current. The method generates essentially the same solutions as the Hallén equation and is efficiently implementable, making use of the numerical tools that we developed in the previous sections for computing the Hallén impedance matrix for different bases,

$$Z_{nm} = \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n - z_m - z)B(z)\,dz \tag{21.14.2}$$

For convenience, we write (21.14.1) in the form:

$$(\partial_z^2 + k^2)V(z) = 2kE_{\rm in}(z)\,, \quad V(z) = \frac{j\eta}{2\pi}\int_{-h}^{h} G(z-z')I(z')\,dz' \tag{21.14.3}$$

Evaluating (21.14.3) at the $N = 2M+1$ sample points $z_n = n\Delta$, $-M \le n \le M$, where the spacing Δ is given by $\Delta = 2h/(2M+1)$ in the pulse, NEC, and delta-function bases, and $\Delta = h/M$ in the triangular basis, we have:

$$(\partial_z^2 + k^2)V(z_n) = 2kE_{\rm in}(z_n)\,, \quad V(z_n) = \frac{j\eta}{2\pi}\int_{-h}^{h} G(z_n - z')I(z')\,dz' \tag{21.14.4}$$

Following [1226], we replace the second derivative in z by the finite difference:

$$\partial_z^2 V(z_n) \simeq \frac{V(z_{n+1}) - 2V(z_n) + V(z_{n-1})}{\Delta^2}$$

Denoting $V_n = V(z_n)$ and $E_n = E_{\rm in}(z_n)$, Eq. (21.14.4) becomes:

$$\frac{V_{n+1} - 2V_n + V_{n-1}}{\Delta^2} + k^2 V_n = 2kE_n$$

which can be re-written as:

$$V_{n+1} - 2\alpha V_n + V_{n-1} = E_n d \tag{21.14.5}$$

where we defined $\alpha = 1 - k^2\Delta^2/2$ and $d = 2k\Delta^2$, and we must restrict n to the range $-(M-1) \le n \le M-1$. Eq. (21.14.5) may be written in the following tridiagonal matrix form that displays *all* the samples V_n and E_n, $-M \le n \le M$, shown here for $M = 3$:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2\alpha & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2\alpha & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2\alpha & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2\alpha & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2\alpha & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_{-3} \\ V_{-2} \\ V_{-1} \\ V_0 \\ V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_{-3} \\ E_{-2} \\ E_{-1} \\ E_0 \\ E_1 \\ E_2 \\ E_3 \end{bmatrix} d \tag{21.14.6}$$

where the top and bottom rows of zeros are redundant and have been added to make these matrices square, both having rank $(N-2)$. We may write (21.14.6) compactly as:

$$\mathcal{A}\boldsymbol{V} = \mathcal{Q}\boldsymbol{E}d \tag{21.14.7}$$

We note that $\mathcal{Q}$ is a projection matrix and so is its complement $\mathcal{P} = I - \mathcal{Q}$,† which enforces the end-point conditions $I_{\pm M} = 0$:

$$\mathcal{P}\boldsymbol{I} = (I - \mathcal{Q})\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I_{-3} \\ I_{-2} \\ I_{-1} \\ I_0 \\ I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} I_{-3} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ I_3 \end{bmatrix} = 0 \tag{21.14.8}$$

The samples V_n can be represented in terms of the Hallén impedance matrix and current samples I_n with respect to a particular basis $B(z)$. Inserting the expansion,

$$I(z) = \sum_{m=-M}^{M} I_m B(z - z_m)$$

into (21.14.4), we have:

$$V_n = \sum_{m=-M}^{M} I_m \frac{j\eta}{2\pi} \int_{-h}^{h} G(z_n - z_m - z) B(z)\,dz = \sum_{m=-M}^{M} Z_{nm} I_m$$

with Z_{nm} given by (21.14.2). Vectorially, we may write

$$\boldsymbol{V} = Z\boldsymbol{I}$$

For the NEC basis, we have instead:

$$I(z) = \sum_{m=-M}^{M} b_m B(z - z_m)\,, \quad \text{and,}$$

$$V_n = \sum_{m=-M}^{M} b_m \frac{j\eta}{2\pi} \int_{-h}^{h} G(z_n - z_m - z) B(z)\,dz = \sum_{m=-M}^{M} Z_{nm} b_m\,, \quad \text{or,}$$

---

†here, I is the identity matrix.

$$\boldsymbol{V} = \mathcal{Z}\boldsymbol{b}\,, \quad \text{with} \quad \boldsymbol{I} = \mathcal{B}\boldsymbol{b} \quad \Rightarrow \quad \boldsymbol{V} = \mathcal{Z}\mathcal{B}^{-1}\boldsymbol{I} \tag{21.14.9}$$

with $\mathcal{B}$ defined as in (21.12.11). Combining with (21.14.7), we obtain:

$$\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}\boldsymbol{I} = \mathcal{Q}\boldsymbol{E}d \tag{21.14.10}$$

Using (21.14.8) and the idempotent property $\mathcal{Q}^2 = \mathcal{Q}$, we have $\boldsymbol{I} = \mathcal{Q}\boldsymbol{I} = \mathcal{Q}^2\boldsymbol{I}$. Making this replacement into (21.14.10), we obtain:

$$\boxed{(\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}\mathcal{Q})\,(\mathcal{Q}\boldsymbol{I}) = (\mathcal{Q}\boldsymbol{E})\,d} \tag{21.14.11}$$

This is the discretized Pocklington equation for the current. The operation $\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}\mathcal{Q}$ simply extracts the middle portion of the matrix $\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}$, and $\mathcal{Q}\boldsymbol{I}$ and $\mathcal{Q}\boldsymbol{E}$ extract the middle portions of $\boldsymbol{I}$ and $\boldsymbol{E}$. Indeed, we have

$$\mathcal{Q}\boldsymbol{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_{-3} \\ E_{-2} \\ E_{-1} \\ E_0 \\ E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} 0 \\ E_{-2} \\ E_{-1} \\ E_0 \\ E_1 \\ E_2 \\ 0 \end{bmatrix} \equiv \left[\begin{array}{c} 0 \\ \hline \bar{\boldsymbol{E}} \\ \hline 0 \end{array}\right]$$

and similarly, $\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}$ and $\mathcal{A}\mathcal{Z}\mathcal{B}^{-1}\mathcal{Q}$ have the structures:

$$\mathcal{A}\mathcal{Z}\mathcal{B}^{-1} = \left[\begin{array}{c|c|c} 0 & \mathbf{0}^T & 0 \\ \hline \mathbf{a} & \bar{Z} & \mathbf{b} \\ \hline 0 & \mathbf{0}^T & 0 \end{array}\right] \quad \Rightarrow \quad \mathcal{A}\mathcal{Z}\mathcal{B}^{-1}\mathcal{Q} = \left[\begin{array}{c|c|c} 0 & \mathbf{0}^T & 0 \\ \hline \mathbf{0} & \bar{Z} & \mathbf{0} \\ \hline 0 & \mathbf{0}^T & 0 \end{array}\right]$$

where $\bar{Z}$ is an $(N-2)\times(N-2)$ non-singular matrix. Thus, (21.14.11) is equivalent to:

$$\left[\begin{array}{c|c|c} 0 & \mathbf{0}^T & 0 \\ \hline \mathbf{0} & \bar{Z} & \mathbf{0} \\ \hline 0 & \mathbf{0}^T & 0 \end{array}\right] \left[\begin{array}{c} 0 \\ \hline \bar{\boldsymbol{I}} \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \bar{\boldsymbol{E}} \\ \hline 0 \end{array}\right] d \quad \Rightarrow \quad \bar{Z}\bar{\boldsymbol{I}} = \bar{\boldsymbol{E}}d \tag{21.14.12}$$

with solution:

$$\boxed{\bar{\boldsymbol{I}} = \bar{Z}^{-1}\bar{\boldsymbol{E}}d}\,, \quad \boldsymbol{I} = \mathcal{Q}\boldsymbol{I} = \left[\begin{array}{c} 0 \\ \hline \bar{\boldsymbol{I}} \\ \hline 0 \end{array}\right] \tag{21.14.13}$$

The MATLAB function `pfield` implements the above solution procedure:

```
[I,z,cnd] = pfield(L,a,E,ker,basis);     solve Pocklington's equation
```

where I, E, z are the vectors of sampled values I_n, E_n, $z_n = n\Delta$, the quantity `cnd` is the condition number of the Pocklington impedance matrix $\bar{Z}$, and the string constants `ker` and `basis` take the possible values:

```
ker = 'e', 'a',                 % exact or approximate kernel
basis = 'p', 't', 'n', 'd',     % pulse, triangular, NEC, or delta-function basis
```

The condition number of the Pocklington matrix $\bar{Z}$ can be one or more orders of magnitude larger than that of the Hallén matrix $\mathcal{Z}$. Yet, the computed currents by the functions `pfield` and `hfield` are remarkably close to each other and virtually indistinguishable over a wide range of the parameters L, a, M.

The large condition number is due to the Helmholtz operator $(\partial_z^2 + k^2)$—represented in discrete form by the matrix $\mathcal{A}$—which causes the Pocklington kernel to become less well-behaved than the Hallén kernel. This has been a primary concern in all the numerical methods for solving Pocklington's equation.

Fig. 21.14.1 shows the currents computed by the Pocklington and Hallén methods induced on a half-wave dipole antenna of radius $a = 0.005\lambda$ by a plane wave incident from broadside (left graph) and by a delta-gap input (right graph).

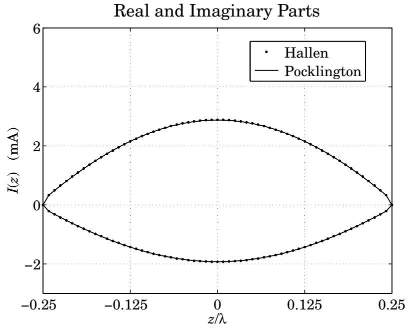


Fig. 21.14.1 Comparison of Pocklington and Hallén equation solutions.

The top curves represent the real parts of the currents and the bottom ones, the imaginary parts. We used a triangular basis with the exact kernel and $M = 30$ segments. The condition numbers of the impedance matrices $\mathcal{Z}$ and $\bar{Z}$ were 17 and 377, respectively. The following MATLAB code illustrates the computation:

```
L = 0.5; a = 0.005; M = 30; ker = 'e'; basis = 't';
if basis=='t', D = L/(2*M); else D = L/(2*M+1); end

E = zeros(2*M+1,1); E(M+1) = 1/D;               % delta-gap input
% E = ones(2*M+1,1);                            % plane-wave input

[If,zf,cf] = hfield(L,a,E,ker,basis);           % Hallén
[Ip,zp,cp] = pfield(L,a,E,ker,basis);           % Pocklington

plot(zf,real(If),'.', zp, real(Ip), '-', zf,imag(If),'.', zp, imag(Ip), '-');
```

To see the dramatic difference between the Hallén and Pocklington impedance matrices $\mathcal{Z}$ and $\bar{Z}$, Fig. 21.14.2 plots the singular values of these matrices (normalized to their maximum value) for the case of a half-wave dipole of radius $a = 0.005\lambda$ using a triangular basis with $M = 100$ and the exact kernel. The computed condition numbers for $\mathcal{Z}$ and $\bar{Z}$ were 59 and 1185, respectively.

The numerical solution of Pocklington's equation has similar properties as the Hallén case, such as, the triangular basis converging faster than the other bases with increasing M, and the use of the approximate kernel causing oscillations at the end-points of the antenna (and at the center for delta-gap input.)

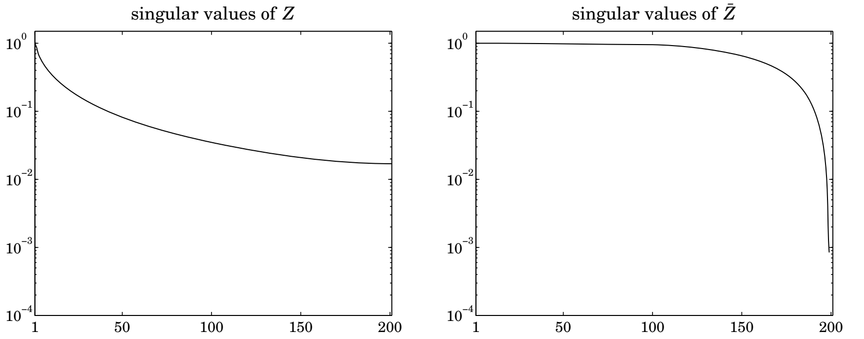


Fig. 21.14.2 Singular values of the impedance matrices $\mathcal{Z}$ and $\bar{Z}$ for $N=2M+1=201$.

This chapter dealt with the currents on a single linear antenna. The case of several antennas forming an array and interacting with each other is treated in Chap. 22.

Hallén's and Pocklington's integral equations generalize into a system of several coupled integral equations for the currents on the antennas. We solve the coupled Hallén equations in the case of delta-gap center-driven antennas. The linearity of the equations allows us to collect them together into a block matrix system from which the currents on each antenna can be obtained.

One simplification arises in the case of an array of identical antennas. Then, the block linear system can be wrapped in half much like it was done in Sec. 21.9, thus, reducing the computational cost.

The case of an array of non-identical antennas is also considered and we obtain solutions for Yagi-Uda arrays with parasitic reflector and director antennas.

21.15 Problems

21.1 Plot the approximations of $\sin kR/kR$ given in Eqs. (21.6.1) and (21.6.18) versus R in the range $R \le 2\lambda$ and verify their validity. Prove the identity:

$$\frac{\sin kR}{kR} = \frac{\sin(kR/8)}{kR/8}\cos(kR/8)\cos(kR/4)\cos(kR/2)$$

which leads to the approximation:

$$\frac{\sin kR}{kR} \simeq \cos(kR/8)\cos(kR/4)\cos(kR/2)$$

Determine the range of applicability of this approximation and plot it together with the previous two cases. Show that it leads to a six-term sinusoidal fit for the current. What are the current basis functions $I_i(z)$, $i = 1, 2, \dots, 6$, in this case?

21.2 *Computer Experiment—Solving Hallén's Equation.* Consider a dipole antenna of length $l = 0.5\lambda$ and radius $a = 0.005\lambda$.

a. For each of the values $M = 20, 50, 100, 200$, solve Hallén's equation for a delta-gap input with voltage $V_0 = 1$ volt using both the exact and the approximate kernels. Plot the real and imaginary parts of the current $I_m = I(z_m)$ versus z_m over the right half of the antenna, that is, $0 \le z_m \le h$, where $h = l/2$.

b. Fit to the computed current samples I_m of the exact kernel to King's three-term approximation. Then, place the fitted points on the same graphs as in part (a). Discuss how well or not the three-term approximation fits the exact-kernel and the approximate-kernel current.

Repeat by using a two-term approximation. Discuss how well or not the two-term approximation fits the exact-kernel and the approximate-kernel current.

c. To illustrate the logarithmic singularity near $z = 0$, evaluate the limiting expression at the points z_m, $m = 1, 2, \dots, M$, for $M = 200$ (the point $z_0 = 0$ is to be skipped):

$$I_{\log}(z_m) = -j\,\frac{4kaV_0}{\eta}\,\ln(k|z_m|) + \text{const.}$$

Adjust the constant so that this expression agrees with the exact-kernel current at the point z_1, that is, $I_{\log}(z_1) = I_1$. Then, plot the imaginary parts of I_m and $I_{\log}(z_m)$ versus z_m.

d. Repeat parts (a–c) for the antenna radius $a = 0.001$ and then for $a = 0.008$. Discuss the effect of changing the radius on the quality of the solution, both for the exact and the approximate kernel cases.

e. Repeat parts (a–d) for the antenna length $l = 1.0\lambda$. Comment on the success of the exact versus approximate kernel calculations versus the parameters l, a, M.

f. For each value of M and current solution I_m, $-M \le m \le M$, the input impedance of the antenna can be calculated from the center sample I_0, that is, $Z_0 = V_0/I_0$. Similarly, the input admittance is:

$$Y_0 = \frac{1}{Z_0} = \frac{I_0}{V_0} = G_0 + jB_0$$

where G_0, B_0 are its real and imaginary parts, that is, the input conductance and susceptance.

For each of the values $M = 1, 2, \dots, 100$, calculate the corresponding conductance and susceptance, $G_0(M), B_0(M)$, using the exact and the approximate kernels and plot them versus M. Use the length and radius $l = 0.5\lambda$ and $a = 0.005\lambda$.

This is a time-consuming question. It requires that you solve the Hallén equation for each value of M for the exact and approximate kernels and pick the center value I_0. Discuss the convergence properties of the exact versus the approximate kernel calculation.

22

Coupled Antennas

22.1 Near Fields of Linear Antennas

In calculating mutual coupling effects between closely-spaced linear antennas, we need to know the fields produced by an antenna at near distances. The fields generated by a thin wire antenna with current $I(z)$ were worked out in Sec. 14.4.

We summarize these results here. All field components can be obtained from the knowledge of the z-component of the magnetic vector potential $A_z(z,\rho)$:

$$\boxed{A_z(z,\rho)=\frac{\mu}{4\pi}\int_{-h}^{h} I(z')\,\frac{e^{-jkR}}{R}\,dz'}\,,\qquad R=\sqrt{\rho^2+(z-z')^2} \tag{22.1.1}$$

where h is the half-length of the antenna, $h = l/2$, and the geometry is shown in Fig. 22.1.1. We have used the approximate thin-wire kernel because it differs little from the exact kernel for distances $\rho > a$ (typically, when $\rho \gtrsim 5a$.)

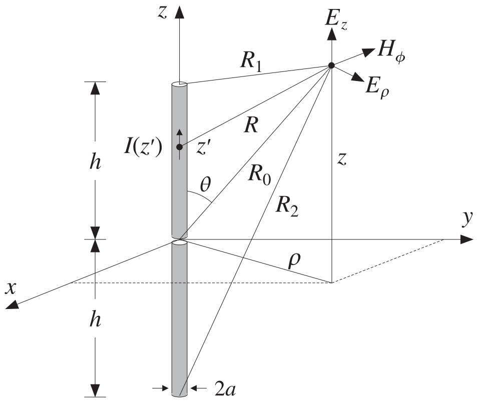


Fig. 22.1.1 Fields of a thin wire antenna.

Then, the non-zero field components E_z, E_ρ, H_ϕ can be constructed from the two alternative sets of formulas:

$$\boxed{\begin{aligned} j\omega\mu\epsilon E_z &= \partial_z^2 A_z + k^2 A_z \\ j\omega\mu\epsilon E_\rho &= \partial_\rho \partial_z A_z \\ \mu H_\phi &= -\partial_\rho A_z \end{aligned}}\,, \qquad \boxed{\begin{aligned} j\omega\mu\epsilon E_z &= \partial_z^2 A_z + k^2 A_z \\ \partial_\rho(\rho H_\phi) &= j\omega\epsilon\,\rho E_z \\ j\omega\epsilon E_\rho &= -\partial_z H_\phi \end{aligned}} \tag{22.1.2}$$

As a first approximation, we will assume that the current $I(z)$ is sinusoidal. This is justified only when the antenna length is near half a wavelength $\lambda/2$. Most coupled antenna arrays that are used in practice, such as Yagi-Uda, satisfy this condition. Thus, we assume that:

$$\boxed{I(z) = I_0\,\frac{\sin\big(k(h-|z|)\big)}{\sin kh} = I_m \sin\big(k(h-|z|)\big)} \tag{22.1.3}$$

where we distinguish between the current I_0 at $z = 0$ and the maximum current $I_m = I_0/\sin kh$. For half-wavelength antennas, we have $kh = \pi/2$, $I_0 = I_m$, and the current becomes $I(z) = I_0\cos kz$.

In principle, one could insert Eq. (22.1.3) into (22.1.1) and perform the required integrations to get A_z. However, for the purpose of determining the fields, this is not necessary. Combining (22.1.1) and (22.1.2), we obtain Pocklington's equation:

$$j\omega\mu\epsilon E_z = \partial_z^2 A_z + k^2 A_z = \frac{\mu}{4\pi}\int_{-h}^{h} I(z')\,(\partial_{z'}^2 + k^2)\,G(z-z')\,dz' \tag{22.1.4}$$

where we denoted $G(z-z') = e^{-jkR}/R$ and replaced ∂_z^2 by $\partial_{z'}^2$. Next, we use the differential identity:

$$I(\partial_{z'}^2 + k^2)G - G(\partial_{z'}^2 + k^2)I = \partial_{z'}\big[I\partial_{z'}G - G\partial_{z'}I\big]$$

Because of the assumed form (22.1.3), $I(z')$ satisfies the Helmholtz equation, $(\partial_{z'}^2 + k^2)I(z') = 0$, and therefore, the integrand of (22.1.4) becomes a complete derivative:

$$I(z')\,(\partial_{z'}^2 + k^2)\,G(z-z') = \partial_{z'}\big[I(z')\,\partial_{z'}G(z-z') - G(z-z')\,\partial_{z'}I(z')\big] \tag{22.1.5}$$

Integrating the first term, we obtain:

$$\int_{-h}^{h}\partial_{z'}\big[I(z')\,\partial_{z'}G(z-z')\big]dz' = I(h)\,\partial_{z'}G(z-h) - I(-h)\,\partial_{z'}G(z+h) = 0$$

where we used the end-conditions $I(h) = I(-h) = 0$. The second term in (22.1.5) is a little trickier because $\partial_{z'}I(z')$ is discontinuous at $z = 0$. Splitting the integration range, we obtain:

$$\begin{aligned} \int_{-h}^{h}\partial_{z'}\big[G(z-z')\,\partial_{z'}I(z')\big]dz' &= \left(\int_{-h}^{0} + \int_{0}^{h}\right)\partial_{z'}\big[G(z-z')\,\partial_{z'}I(z')\big]dz' \\ &= \big[G(z)I'(0-) - G(z+h)I'(-h)\big] + \big[G(z-h)I'(h) - G(z)I'(0+)\big] \\ &= kI_m\big[2\cos kh\,G(z) - G(z-h) - G(z+h)\big] \end{aligned}$$

where we used $I'(0\pm) = \mp kI_m\cos kh$ and $I'(\pm h) = \mp kI_m$. Inserting this result into Eq. (22.1.4) and rearranging some constants, we find:

$$E_z(z,\rho) = -\frac{j\eta I_m}{4\pi}\big[G(z-h) + G(z+h) - 2\cos kh\,G(z)\big] \tag{22.1.6}$$

The quantities $G(z-h), G(z+h), G(z)$ can be written conveniently as follows:

$$\begin{aligned} G(z) &= \frac{e^{-jkR_0}}{R_0}\,, & R_0 &= \sqrt{\rho^2 + z^2} \\ G(z-h) &= \frac{e^{-jkR_1}}{R_1}\,, & R_1 &= \sqrt{\rho^2 + (z-h)^2} \\ G(z+h) &= \frac{e^{-jkR_2}}{R_2}\,, & R_2 &= \sqrt{\rho^2 + (z+h)^2} \end{aligned} \tag{22.1.7}$$

where R_0, R_1, R_2 are recognized to be the distances from the center and the two ends of the antenna to the observation point, as shown in Fig. 22.1.1. Thus, we can write:

$$\boxed{E_z(z,\rho) = -\frac{j\eta I_m}{4\pi}\left[\frac{e^{-jkR_1}}{R_1} + \frac{e^{-jkR_2}}{R_2} - 2\cos kh\,\frac{e^{-jkR_0}}{R_0}\right]} \tag{22.1.8}$$

Next, we determine H_ϕ from Ampère's law in (22.1.2) by noting that ρE_z is a complete derivative with respect to ρ. Indeed, for any of the quantities R, we have:

$$\partial_\rho(e^{-jkR}) = -jk(\partial_\rho R)\,e^{-jkR} = -jk\rho\frac{e^{-jkR}}{R} \quad\Rightarrow\quad \frac{e^{-jkR}}{R} = -\frac{1}{jk\rho}\,\partial_\rho(e^{-jkR})$$

Applying this result to all three terms of Eq. (22.1.8), we have:

$$\rho E_z(z,\rho) = -\frac{j\eta I_m}{4\pi}\,\frac{1}{-jk}\,\partial_\rho\big[e^{-jkR_1} + e^{-jkR_2} - 2\cos kh\,e^{-jkR_0}\big]$$

Inserting this into Ampère's law, $\partial_\rho(\rho H_\phi) = j\omega\epsilon\,\rho E_z$, and rearranging some constants, we find:

$$\partial_\rho(\rho H_\phi) = \frac{jI_m}{4\pi}\,\partial_\rho\big[e^{-jkR_1} + e^{-jkR_2} - 2\cos kh\,e^{-jkR_0}\big]$$

which can be integrated trivially, giving:

$$\boxed{H_\phi(z,\rho) = \frac{jI_m}{4\pi\rho}\big[e^{-jkR_1} + e^{-jkR_2} - 2\cos kh\,e^{-jkR_0}\big]} \tag{22.1.9}$$

A possible integration constant in ρ is dropped because the field must vanish when its source vanishes, that is, when $I_m = 0$. Finally, we obtain E_ρ from Faraday's law in (22.1.2). Noting the differentiation property:

$$\partial_z(e^{-jkR}) = -jk\,\frac{z}{R}\,e^{-jkR}\,, \qquad R = \sqrt{\rho^2 + z^2}$$

we obtain from $j\omega\epsilon E_\rho = -\partial_z H_\phi$:

$$\boxed{E_\rho(z,\rho) = \frac{j\eta I_m}{4\pi\rho}\left[\frac{z-h}{R_1}\,e^{-jkR_1} + \frac{z+h}{R_2}\,e^{-jkR_2} - 2\cos kh\,\frac{z}{R_0}\,e^{-jkR_0}\right]} \tag{22.1.10}$$

It is worth also to verify that the exact expressions for the fields give correctly the radiation fields that were derived in Sec. 16.3. At large distances, we can make the approximations:

$$R_0 = r, \quad R_1 = r - h\cos\theta, \quad R_2 = r + h\cos\theta$$

where r is the radial distance and θ the polar angle. Replacing $\rho = r\sin\theta$, the magnetic field (22.1.9) becomes approximately:

$$H_\phi(r,\theta) = \frac{jI_m}{4\pi r\sin\theta}\left[e^{-jk(r-h\cos\theta)} + e^{-jk(r+h\cos\theta)} - 2\cos kh\,e^{-jkr}\right]$$

which simplifies into:

$$H_\phi(r,\theta) = \frac{jI_m e^{-jkr}}{2\pi r}\,\frac{\cos(kh\cos\theta) - \cos kh}{\sin\theta} \tag{22.1.11}$$

This agrees with the results of Sec. 16.3.

22.2 Self and Mutual Impedance

The mutual coupling between antennas cannot be ignored if the antennas are near each other. The *mutual impedance* is a measure of such proximity effects [2,1277–1289].

Consider two parallel center-driven linear dipoles, as shown in Fig. 22.2.1. Their distance along the x-direction is d and their centers are offset by b along the z-direction.

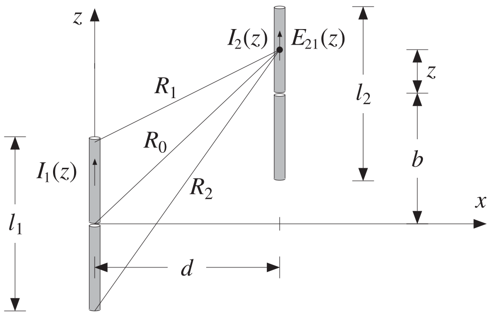


Fig. 22.2.1 Parallel linear dipoles.

If antenna-1 is driven and antenna-2 is open-circuited, the near field generated by the current on antenna-1 will cause an open-circuit voltage, say $V_{21,\text{oc}}$ on antenna-2. The mutual impedance of antenna-2 due to antenna-1 is defined to be:

$$Z_{21} = \frac{V_{21,\text{oc}}}{I_1} \tag{22.2.1}$$

where I_1 is the input current on antenna-1. Reciprocity implies that $Z_{12} = Z_{21}$. More generally, if both antennas are driven, then, the relationship of the driving voltages to the input currents is given by:

$$\boxed{\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 \\ V_2 &= Z_{21}I_1 + Z_{22}I_2 \end{aligned}} \tag{22.2.2}$$

The quantities Z_{11}, Z_{22} are the *self impedances* of the two antennas and are approximately equal to the input impedances of the isolated antennas, that is, when the other antenna is absent. If antenna-2 is open-circuited, so that $I_2 = 0$, then the second of Eqs. (22.2.2) gives (22.2.1).

In order to derive convenient expressions that allow the calculation of the mutual and self impedances, we use the reciprocity result given in Eq. (21.5.6) for the short-circuit current and open-circuit voltage induced on a receiving antenna in the presence of an incident field.

If antenna-2 is open-circuited and the z-component of the electric field generated by antenna-1 and incident on antenna-2 is $E_{21}(z)$, then according to Eq. (21.5.6), the induced open-circuit voltage will be:

$$V_{21,\text{oc}} = -\frac{1}{I_2}\int_{-h_2}^{h_2} E_{21}(z)I_2(z)\,dz \tag{22.2.3}$$

where $h_2 = l_2/2$, and $I_2(z)$, $I_2 = I_2(0)$ are the current and input current on antenna-2 when it is transmitting. It follows from definition (22.2.1) that:

$$\boxed{Z_{21} = \frac{V_{21,\text{oc}}}{I_1} = -\frac{1}{I_1 I_2}\int_{-h_2}^{h_2} E_{21}(z)I_2(z)\,dz} \tag{22.2.4}$$

Assuming that the currents are sinusoidal,

$$\begin{aligned} I_1(z) &= I_1\,\frac{\sin\big(k(h_1 - |z|)\big)}{\sin kh_1} = I_{m1}\sin\big(k(h_1 - |z|)\big) \\ I_2(z) &= I_2\,\frac{\sin\big(k(h_2 - |z|)\big)}{\sin kh_2} = I_{m2}\sin\big(k(h_2 - |z|)\big) \end{aligned}$$

then, according to Eq. (22.1.8) the electric field $E_{21}(z)$ along antenna-2 will be:

$$\boxed{E_z(z) = -\frac{j\eta I_{m1}}{4\pi}\left[\frac{e^{-jkR_1}}{R_1} + \frac{e^{-jkR_2}}{R_2} - 2\cos kh_1\,\frac{e^{-jkR_0}}{R_0}\right]} \tag{22.2.5}$$

where $-h_2 \le z \le h_2$, and R_1, R_2, R_0 are defined in Fig. 22.2.1:

$$\begin{aligned} R_0 &= \sqrt{d^2 + (z+b)^2} \\ R_1 &= \sqrt{d^2 + (z+b-h_1)^2} \\ R_2 &= \sqrt{d^2 + (z+b+h_1)^2} \end{aligned} \tag{22.2.6}$$

Inserting Eq. (22.2.5) into (22.2.4) and rearranging some constants, we find the final expression for the mutual impedance Z_{21}:

$$\boxed{Z_{21} = \frac{j\eta}{4\pi \sin kh_1\, \sin kh_2}\int_{-h_2}^{h_2} F(z)\,dz} \tag{22.2.7}$$

$$\boxed{F(z) = \left[\frac{e^{-jkR_1}}{R_1} + \frac{e^{-jkR_2}}{R_2} - 2\cos kh_1\, \frac{e^{-jkR_0}}{R_0}\right]\sin\big(k(h_2 - |z|)\big)} \tag{22.2.8}$$

This is the mutual impedance referred to the *input terminals* of the antennas. If one or both of the antennas have lengths that are multiples of λ, then one or both of the denominator factors $\sin kh_1$, $\sin kh_2$ will vanish resulting in an infinite value for the mutual impedance.

This limitation is caused by the sinusoidal current assumption. We saw in Chap. 21 that the actual input currents are not zero in a real antenna. On the other hand, in most applications of Eq. (22.2.7) the lengths differ slightly from half-wavelength for which the sinusoidal approximation is good.

The definition (22.2.4) can also be referred to the *maximum currents* by normalizing by the factor $I_{m1}I_{m2}$, instead of I_1I_2. In this case, the mutual impedance is $Z_{21m} = Z_{21}\sin kh_1\, \sin kh_2$, that is,

$$Z_{21m} = \frac{j\eta}{4\pi}\int_{-h_2}^{h_2} F(z)\,dz \tag{22.2.9}$$

The *self-impedance* of a single antenna can be calculated also by the same formula (22.2.7). Evaluating the near-field on the surface of the single antenna, that is, at $d = a$, where a is the antenna radius, and setting $h_2 = h_1$ and $b = 0$ in Eq. (22.2.6), we find:

$$\boxed{Z_{11} = -\frac{1}{I_1^2}\int_{-h_1}^{h_1} E_{11}(z)I_1(z)\,dz = \frac{j\eta}{4\pi \sin^2 kh_1}\int_{-h_1}^{h_1} F(z)\,dz} \tag{22.2.10}$$

$$\boxed{F(z) = \left[\frac{e^{-jkR_1}}{R_1} + \frac{e^{-jkR_2}}{R_2} - 2\cos kh_1\, \frac{e^{-jkR_0}}{R_0}\right]\sin\big(k(h_1 - |z|)\big)} \tag{22.2.11}$$

$$R_0 = \sqrt{a^2 + z^2}\,,\quad R_1 = \sqrt{a^2 + (z - h_1)^2}\,,\quad R_2 = \sqrt{a^2 + (z + h_1)^2} \tag{22.2.12}$$

The MATLAB function `imped` implements Eq. (22.2.7), as well as (22.2.10). It returns both Z_{21} and Z_{21m} and has usage:

```
[Z21,Z21m] = imped(L2,L1,d,b)      % mutual impedance of dipole 2 due to dipole 1
[Z21,Z21m] = imped(L2,L1,d)        % b = 0, side-by-side arrangement
    [Z,Zm] = imped(L,a)            % self impedance
```

where all the lengths are in units of λ. The function uses 16-point Gauss-Legendre integration, implemented with the help of the function `quadr`, to perform the integral in Eq. (22.2.7).

In evaluating the self impedance of an antenna with a small radius, the integrand $F(z)$ varies rapidly around $z = 0$. To maintain accuracy in the integration, we split the integration interval into three subintervals, as we mentioned in Sec. 21.10.

Example 22.2.1: Because the function `imped` uses an even length (that is, 16) for the Gauss-Legendre integration, the integrand $F(z)$ is never evaluated at $z = 0$, even if the antenna radius is zero. This allows us to estimate the self-impedance of an infinitely thin half-wavelength antenna by setting $L = 0.5$ and $a = 0$:

$$Z = \text{imped}(0.5, 0) = 73.0790 + 42.5151j\ \ \Omega$$

Similarly, for radii $a = 0.001\lambda$ and 0.005λ, we find:

$$Z = \text{imped}(0.5, 0.001) = 73.0784 + 42.2107j\ \ \Omega$$
$$Z = \text{imped}(0.5, 0.005) = 73.0642 + 40.6319j\ \ \Omega$$

A resonant antenna is obtained by adjusting the length L such that the reactance part of Z becomes zero. The resonant length depends on the antenna radius. For zero radius, this length is $L = 0.48574823$ and the corresponding impedance, $Z = 67.1843\ \ \Omega$. □

Example 22.2.2: Consider two identical parallel half-wavelength dipoles in side-by-side arrangement separated by distance d. The antenna radius is $a = 0.001$ and therefore, its self impedance is as in the previous example. If antenna-1 is driven and antenna-2 is parasitic, that is, short-circuited, then Eq. (22.2.2) gives:

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 \\ 0 &= Z_{21}I_1 + Z_{22}I_2 \end{aligned}$$

Solving the second for the parasitic current $I_2 = -I_1 Z_{21}/Z_{22}$ and substituting in the first, we obtain driving-point impedance of the first antenna:

$$Z_{\text{in}} = \frac{V_1}{I_1} = Z_{11} - \frac{Z_{12}Z_{21}}{Z_{22}} = Z_{11}\left(1 - \frac{Z_{21}^2}{Z_{11}^2}\right)$$

where we used $Z_{12} = Z_{21}$ and $Z_{22} = Z_{11}$. The ratio Z_{21}^2/Z_{11}^2 quantifies the effect of the coupling and the deviation of Z_{in} from Z_{11}. For example, we find the values:

d	0.125λ	0.25λ	0.50λ	0.75λ	1.00λ
$\lvert Z_{21}/Z_{11}\rvert^2$	0.58	0.35	0.15	0.08	0.05

Thus, the ratio decreases rapidly with increasing distance d. Fig. 22.2.2 shows a plot of Z_{21} versus distance d. □

For separations d that are much larger than the antenna lengths, the impedance Z_{21} falls like $1/d$. Indeed, it follows from Eq. (22.2.6) that for large d, all three distances R_0, R_1, R_2 become equal to d. Therefore, (22.2.8) tends to:

$$F(z) \to \frac{e^{-jkd}}{d}\big(2 - 2\cos kh_1\big)\sin\big(k(h_2 - |z|)\big)$$

which, when inserted into (22.2.7), gives the asymptotic form:

$$Z_{21} \to \frac{j\eta(1 - \cos kh_1)(1 - \cos kh_2)}{\pi \sin kh_1 \sin kh_2}\,\frac{e^{-jkd}}{kd}\,,\quad \text{for large } d \tag{22.2.13}$$

The envelope of this asymptotic form was superimposed on the graph of Fig. 22.2.2. The oscillatory behavior of Z_{21} with distance is essentially due to the factor e^{-jkd}.

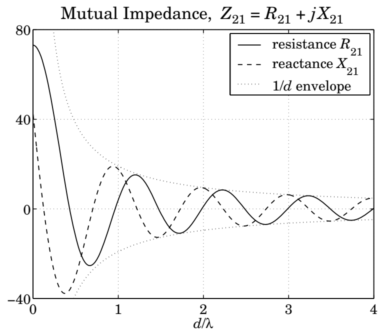


Fig. 22.2.2 Mutual impedance between identical half-wave dipoles vs. separation.

An alternative computation method of the mutual impedance is to reduce the integrals (22.2.7) to the exponential integral $E_1(z)$ defined in Appendix F, taking advantage of MATLAB's built-in function `expint`.

By folding the integration range $[-h_1, h_1]$ in half and writing $\sin\big(k(h_2 - |z|)\big)$ as a sum of exponentials, Eq. (22.2.7) can be reduced to a sum of terms of the form:

$$G(z_0, s) = \int_0^{h_1} \frac{e^{-jkR}}{R}\, e^{-jksz}\, dz\,, \quad R = \sqrt{d^2 + (z - z_0)^2}\,, \quad s = \pm 1 \tag{22.2.14}$$

which can be evaluated in terms of $E_1(z)$ as:

$$G(z_0, s) = s e^{-jksz_0}\big[E_1(ju_0) - E_1(ju_1)\big] \tag{22.2.15}$$

with

$$u_0 = k\left[\sqrt{d^2 + z_0^2} - sz_0\right]$$
$$u_1 = k\left[\sqrt{d^2 + (h_1 - z_0)^2} + s(h_1 - z_0)\right]$$

Indeed, the integral in (22.2.7) can be written as a linear combination of 10 such terms:

$$\int_{-h_1}^{h_1} F(z)\,dz = \sum_{i=1}^{10} c_i\, G(z_i, s_i) \tag{22.2.16}$$

with the following values of z_i, c_i, and s_i, where $c_1 = e^{jkh_2}/(2j)$:

i	z_i	s_i	c_i	i	z_i	s_i	c_i
1	$h_1 - b$	1	c_1	6	$h_1 - b$	-1	c_1^*
2	$-h_1 + b$	1	c_1	7	$-h_1 + b$	-1	c_1^*
3	$-h_1 - b$	1	c_1	8	$-h_1 - b$	-1	c_1^*
4	$h_1 + b$	1	c_1	9	$h_1 + b$	-1	c_1^*
5	b	1	$-4c_1\cos kh_1$	10	b	-1	$-4c_1^*\cos kh_1$

The MATLAB function `Gi` implements the "Green's function integral" of (22.2.14). The function `imped2`, which is an alternative to `imped`, uses (22.2.16) to calculate (22.2.7).

The input impedance (22.2.10) deserves a closer look. Replacing the exponential integrals in (22.2.16) in terms of their real and imaginary parts,

$$E_1(ju) = -\gamma - \ln u + C_{\text{in}}(u) + j\big(S_i(u) - \frac{\pi}{2}\big)$$

as defined in Eq. (F.27), then (22.2.10) can be expressed in the following form, where we set $Z_{11} = Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}}$, $h_1 = h$, and $l = 2h$:

$$Z_{\text{in}} = R_{\text{in}} + jX_{\text{in}} = \frac{\eta}{2\pi}\,\frac{A + jB}{\sin^2 kh} \tag{22.2.17}$$

With the definitions $l_\pm = \sqrt{a^2 + h^2} \pm h$ and $L_\pm = \sqrt{a^2 + 4h^2} \pm 2h$, we obtain:

$$\begin{aligned} A &= C_{\text{in}}(kl_+) + C_{\text{in}}(kl_-) - 2C_{\text{in}}(ka) \\ &\quad + \frac{1}{2}\cos kl\,\big[2C_{\text{in}}(kl_+) - C_{\text{in}}(kL_+) + 2C_{\text{in}}(kl_-) - C_{\text{in}}(kL_-) - 2C_{\text{in}}(ka)\big] \\ &\quad + \frac{1}{2}\sin kl\,\big[2S_i(kl_-) - S_i(kL_-) + S_i(kL_+) - 2S_i(kl_+)\big] \end{aligned} \tag{22.2.18}$$

$$\begin{aligned} B &= S_i(kl_+) + S_i(kl_-) - 2S_i(ka) \\ &\quad + \frac{1}{2}\cos kl\,\big[2S_i(kl_+) - S_i(kL_+) + 2S_i(kl_-) - S_i(kL_-) - 2S_i(ka)\big] \\ &\quad + \frac{1}{2}\sin kl\left[2C_{\text{in}}(kl_+) - C_{\text{in}}(kL_+) + C_{\text{in}}(kL_-) - 2C_{\text{in}}(kl_-) + 2\ln\left(\frac{aL_+}{l_+^2}\right)\right] \end{aligned} \tag{22.2.19}$$

These expressions simplify substantially if we assume that the radius a is small, as is the case in practice. In particular, assuming that $ka \ll 1$ and $a \ll h$, the quantities $l_\pm$ and $L_\pm$ can be approximated by:

$$\begin{aligned} l_+ &\simeq 2h = l\,, & l_- &= \frac{a^2}{l_+} \simeq \frac{a^2}{l} \\ L_+ &\simeq 4h = 2l\,, & L_- &= \frac{a^2}{L_+} \simeq \frac{a^2}{2l} \end{aligned} \tag{22.2.20}$$

Noting that $S_i(x)$ and $C_{\text{in}}(x)$ vanish at $x = 0$, we may neglect all the terms whose arguments are kl_-, kL_-, or ka, and replace $kl_+ = kl$ and $kL_+ = 2kl$, obtaining:

$$A = C_{\text{in}}(kl) + \frac{1}{2}\cos kl\,\big[2C_{\text{in}}(kl) - C_{\text{in}}(2kl)\big] + \frac{1}{2}\sin kl\,\big[S_i(2kl) - 2S_i(kl)\big] \tag{22.2.21}$$

$$B = S_i(kl) + \frac{1}{2}\cos kl\,\big[2S_i(kl) - S_i(2kl)\big] + \frac{1}{2}\sin kl\left[2C_{\text{in}}(kl) - C_{\text{in}}(2kl) + 2\ln\left(\frac{2a}{l}\right)\right] \tag{22.2.22}$$

We note that A is independent of the radius a and leads to the same expression for the radiation resistance that we found in Sec. 16.3 using Poynting methods.

An additional approximation can be made for the case of a *small dipole*. Assuming that $kh \ll 1$, in addition to $ka \ll 1$ and $a \ll h$, we may expand each of the above terms into a Taylor series in the variable kh using the following Taylor series expansions of the functions $S_i(x)$ and $C_{\rm in}(x)$:

$$S_i(x)\simeq x-\frac{1}{18}x^3+\frac{1}{600}x^5\,,\quad C_{\rm in}(x)\simeq\frac{1}{4}x^2-\frac{1}{96}x^4+\frac{1}{4320}x^6 \tag{22.2.23}$$

Such expansions, lead to the following input impedance $Z=R+jX$ to the lowest non-trivial order in kl:

$$Z_{\rm in}=R_{\rm in}+jX_{\rm in}=\frac{\eta}{2\pi}\left[\frac{1}{12}(kl)^2+j\frac{4(1+L)}{kl}\right]\quad\text{(small dipole)} \tag{22.2.24}$$

where $L=\ln(2a/l)$. The resistance R is identical to that obtained using the Poynting method and assuming a linear approximation to the sinusoidal antenna current, which is justified when $kh\ll 1$:

$$I(z)=I_0\frac{\sin\big(k(h-|z|)\big)}{\sin kh}\simeq I_0\frac{k(h-|z|)}{kh}=I_0\left(1-\frac{|z|}{h}\right) \tag{22.2.25}$$

22.3 Coupled Two-Element Arrays

Next, we consider a more precise justification of Eq. (22.2.2) and generalize it to the case of an arbitrary array of parallel linear antennas. Fig. 22.3.1 shows two z-directed parallel dipoles with centers at locations (x_1,y_1) and (x_2,y_2).

We assume that the dipoles are center-driven by the voltage generators V_1,V_2. Let $I_1(z),I_2(z)$ be the currents induced on the dipoles by the generators and by their mutual interaction, and let h_1,h_2 be the half-lengths of the antennas, and a_1,a_2, their radii. Then, assuming the thin-wire model, the total current density will have only a z-component given by:

$$J_z(x',y',z')=I_1(z')\delta(x'-x_1)\delta(y'-y_1)+I_2(z')\delta(x'-x_2)\delta(y'-y_2) \tag{22.3.1}$$

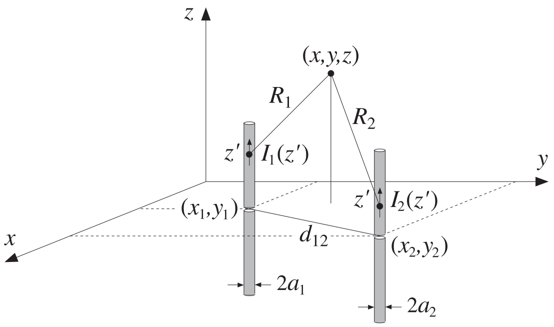


Fig. 22.3.1 Array of two linear antennas.

It follows that the magnetic vector potential will be:

$$A_z(z,\boldsymbol{\rho})=\frac{\mu}{4\pi}\int\frac{e^{-jkR}}{R}J_z(x',y',z')\,dx'dy'dz'\,,\qquad R=|\mathbf{r}-\mathbf{r}'|$$

where $\boldsymbol{\rho}=x\hat{\mathbf{x}}+y\hat{\mathbf{y}}$ is the cylindrical radial vector. Inserting (22.3.1) and performing the x',y' integrations, we obtain:

$$A_z(z,\boldsymbol{\rho})=\frac{\mu}{4\pi}\int_{-h_1}^{h_1}\frac{e^{-jkR_1}}{R_1}I_1(z')\,dz'+\frac{\mu}{4\pi}\int_{-h_2}^{h_2}\frac{e^{-jkR_2}}{R_2}I_2(z')\,dz' \tag{22.3.2}$$

where, as shown in Fig. 22.3.1, R_1,R_2 are the distances from the z' point on each antenna to the (x,y,z) observation point, that is,

$$\begin{aligned}R_1&=\sqrt{(z-z')^2+(x-x_1)^2+(y-y_1)^2}=\sqrt{(z-z')^2+|\boldsymbol{\rho}-\boldsymbol{d}_1|^2}\\R_2&=\sqrt{(z-z')^2+(x-x_2)^2+(y-y_2)^2}=\sqrt{(z-z')^2+|\boldsymbol{\rho}-\boldsymbol{d}_2|^2}\end{aligned} \tag{22.3.3}$$

where $\boldsymbol{d}_1=(x_1,y_1)$ and $\boldsymbol{d}_2=(x_2,y_2)$ are the xy-locations of the antenna centers. The z-component of the electric field generated by the two antenna currents will be:

$$j\omega\epsilon\mu\,E_z(z,\boldsymbol{\rho})=(\partial_z^2+k^2)A_z(z,\boldsymbol{\rho})$$

Working with the rescaled vector potential $V(z,\boldsymbol{\rho})=2jcA_z(z,\boldsymbol{\rho})$, we rewrite:

$$V(z,\boldsymbol{\rho})=\frac{j\eta}{2\pi}\int_{-h_1}^{h_1}\frac{e^{-jkR_1}}{R_1}I_1(z')\,dz'+\frac{j\eta}{2\pi}\int_{-h_2}^{h_2}\frac{e^{-jkR_2}}{R_2}I_2(z')\,dz' \tag{22.3.4}$$

$$(\partial_z^2+k^2)V(z,\boldsymbol{\rho})=-2kE_z(z,\boldsymbol{\rho}) \tag{22.3.5}$$

Denoting by $V_1(z)$ and $V_2(z)$ the values of $V(x,y,z)$ on the surfaces of antenna-1 and antenna-2, we obtain from Eq. (22.3.4):

$$\begin{aligned}V_1(z)&=V_{11}(z)+V_{12}(z)\\V_2(z)&=V_{21}(z)+V_{22}(z)\end{aligned} \tag{22.3.6}$$

The z-components of the electric fields induced on the surfaces of antenna-1 and antenna-2 are obtained by applying Eq. (22.3.5) to each term of (22.3.6):

$$\begin{aligned}E_1(z)&=E_{11}(z)+E_{12}(z)\\E_2(z)&=E_{21}(z)+E_{22}(z)\end{aligned} \tag{22.3.7}$$

where we defined, for $p,q=1,2$:

$$\boxed{\begin{gathered}V_{pq}(z)=\frac{j\eta}{2\pi}\int_{-h_q}^{h_q}G_{pq}(z-z')I_q(z')\,dz'\\(\partial_z^2+k^2)V_{pq}(z)=-2kE_{pq}(z)\end{gathered}} \tag{22.3.8}$$

and the impedance kernels:

$$\boxed{G_{pq}(z-z')=\frac{e^{-jkR_{pq}}}{R_{pq}}}\,,\quad R_{pq}=\sqrt{(z-z')^2+d_{pq}^2} \tag{22.3.9}$$

If $p \neq q$, then d_{pq} is the xy-distance between the antennas, and if $p = q$, it is the radius of the corresponding antenna, that is,

$$\begin{aligned} d_{12} &= d_{21} = |\boldsymbol{d}_1 - \boldsymbol{d}_2| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ d_{11} &= a_1, \quad d_{22} = a_2 \end{aligned} \tag{22.3.10}$$

Thus, $V_{pq}(z)$ and $E_{pq}(z)$ are the vector potential and the z-component of the electric field induced on antenna-p by the current $I_q(z)$ on antenna-q.

To clarify these definitions, Fig. 22.3.2 shows a projected view of Fig. 22.3.1 on the xy plane. The point P with radial vector $\boldsymbol{\rho}$ is the projection of the observation point $(z, \boldsymbol{\rho})$. When P coincides with a point, such as P_2, on the surface of antenna-2 defined by the radial vector $\boldsymbol{\rho}_2$, then the distance $(P_2O_2) = |\boldsymbol{\rho}_2 - \boldsymbol{d}_2|$ will be equal to the antenna radius a_2, regardless of the location of P_2 around the periphery of the antenna.

On the other hand, the distance $(P_2O_1) = |\boldsymbol{\rho}_2 - \boldsymbol{d}_1|$ varies with P_2. However, because the separation d_{12} is typically $d_{12} \gg a_2$, such variation is minor and we may replace $|\boldsymbol{\rho}_2 - \boldsymbol{d}_1|$ by $|\boldsymbol{d}_2 - \boldsymbol{d}_1|$. Thus, in evaluating $V(z, \boldsymbol{\rho}_2)$ on antenna-2, we may use Eq. (22.3.4) with R_1, R_2 defined by:

$$\begin{aligned} R_1 &= \sqrt{(z - z')^2 + |\boldsymbol{d}_2 - \boldsymbol{d}_1|^2} = \sqrt{(z - z')^2 + d_{12}^2} \\ R_2 &= \sqrt{(z - z')^2 + |\boldsymbol{\rho}_2 - \boldsymbol{d}_2|^2} = \sqrt{(z - z')^2 + a_2^2} \end{aligned} \tag{22.3.11}$$

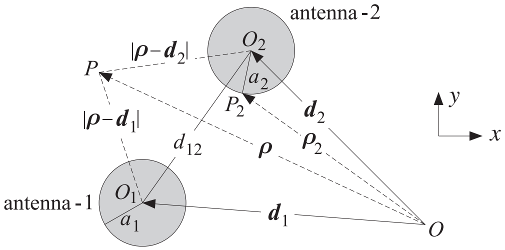


Fig. 22.3.2 Array of two linear antennas.

Now, on the surface of the first antenna, the electric field E_z must cancel the field of the delta-gap generator in order for the total tangential field to vanish, that is, $E_1(z) = -E_{1,\text{in}}(z) = -V_1\delta(z)$. Similarly, on the surface of the second antenna, we must have $E_2(z) = -E_{2,\text{in}}(z) = -V_2\delta(z)$. Then, Eq. (22.3.7) becomes:

$$\begin{aligned} E_{11}(z) + E_{12}(z) = E_1(z) = -V_1\delta(z) \\ E_{21}(z) + E_{22}(z) = E_2(z) = -V_2\delta(z) \end{aligned} \tag{22.3.12}$$

Combining these with the Eq. (22.3.8), we obtain the coupled version of the Hallén-Pocklington equations:

$$\begin{aligned} (\partial_z^2 + k^2)\big[V_{11}(z) + V_{12}(z)\big] &= 2kV_1\delta(z) \\ (\partial_z^2 + k^2)\big[V_{21}(z) + V_{22}(z)\big] &= 2kV_2\delta(z) \end{aligned} \tag{22.3.13}$$

We will solve these numerically in Sec. 22.6. Next, we derive Eq. (22.2.2). According to definitions (22.2.4) and (22.2.10), the mutual impedance between antenna-p and antenna-q can be restated as follows, for $p, q = 1, 2$:

$$\boxed{Z_{pq} = -\frac{1}{I_p I_q}\int_{-h_p}^{h_p} E_{pq}(z) I_p(z)\, dz} \tag{22.3.14}$$

and, more explicitly:

$$\begin{aligned} Z_{11} &= -\frac{1}{I_1 I_1}\int_{-h_1}^{h_1} E_{11}(z) I_1(z)\, dz\,, \quad Z_{12} = -\frac{1}{I_1 I_2}\int_{-h_1}^{h_1} E_{12}(z) I_1(z)\, dz \\ Z_{21} &= -\frac{1}{I_2 I_1}\int_{-h_2}^{h_2} E_{21}(z) I_2(z)\, dz\,, \quad Z_{22} = -\frac{1}{I_2 I_2}\int_{-h_2}^{h_2} E_{22}(z) I_2(z)\, dz \end{aligned}$$

Using these definitions and Eq. (22.3.12), we find:

$$\begin{aligned} Z_{11}I_1 + Z_{12}I_2 &= -\frac{1}{I_1}\int_{-h_1}^{h_1}\big[E_{11}(z) + E_{12}(z)\big] I_1(z)\, dz \\ &= -\frac{1}{I_1}\int_{-h_1}^{h_1}\big[-V_1\delta(z)\big] I_1(z)\, dz = \frac{1}{I_1} V_1 I_1(0) = V_1 \end{aligned}$$

where, by definition, $I_1(0) = I_1$. Similarly, we can show the second of Eq. (22.2.2).

The mutual impedance defined in Eq. (22.3.14) actually satisfies the reciprocity symmetry condition, $Z_{pq} = Z_{qp}$. To write it in a form that shows this condition explicitly, we replace $E_{pq}(z)$ by (22.3.8) and obtain the alternative symmetric form:

$$\boxed{Z_{pq} = \frac{j\eta}{4\pi k}\int_{-h_p}^{h_p}\int_{-h_q}^{h_q} \frac{I_p(z) I_q(z')}{I_p I_q}(\partial_z^2 + k^2) G_{pq}(z - z')\, dz\, dz'} \tag{22.3.15}$$

If we assume that the currents are sinusoidal, that is, for $p = 1, 2$,

$$I_p(z) = I_p \frac{\sin\big(k(h_p - |z|)\big)}{\sin kh_p} \tag{22.3.16}$$

then, in Eq. (22.3.15) the ratios $I_p(z)/I_p$ and hence Z_{pq} become independent of the input currents at the antenna terminals and depend only on the geometry of the antennas.

22.4 Arrays of Parallel Dipoles

The above results on two antennas generalize in a straightforward fashion to several antennas. Fig. 22.4.1 depicts the case of K parallel dipoles in side-by-side arrangement with centers at positions (x_p, y_p), and driving voltages, lengths, half-lengths, and radii, V_p, l_p, h_p, a_p, where $p = 1, 2\dots, K$.

Assuming sinusoidal currents as in Eq. (22.3.16), we define the mutual impedances Z_{pq} by Eq. (22.3.14) or (22.3.15), where p, q take on the values $p, q = 1, 2\dots, K$. The

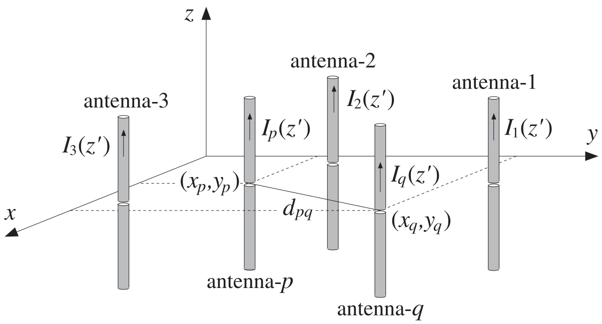


Fig. 22.4.1 Two-dimensional array of parallel dipoles.

Hallén-Pocklington equations (22.3.13) generalize into:

$$\boxed{(\partial_z^2+k^2)\sum_{q=1}^{K}V_{pq}(z)=-2k\sum_{q=1}^{K}E_{pq}(z)=2kV_p\delta(z)}\,,\quad p=1,2,\dots,K \tag{22.4.1}$$

where $V_{pq}(z)$ is defined by Eqs. (22.3.8) and (22.3.9). The mutual distances are:

$$d_{pq}=\begin{cases}\sqrt{(x_p-x_q)^2+(y_p-y_q)^2}\,, & \text{if}\quad p\neq q\\ a_p, & \text{if}\quad p=q\end{cases} \tag{22.4.2}$$

Multiplying Eq. (22.4.1) by $I_p(z)$ and integrating along the length of the pth antenna, and using the mutual impedance definitions (22.3.14), we obtain the generalization of Eq. (22.2.2) to the case of K antennas:

$$\boxed{V_p=\sum_{q=1}^{K}Z_{pq}I_q}\,,\quad p=1,2,\dots,K \tag{22.4.3}$$

where I_q is the input current at the center of the qth antenna. Eq. (22.4.3) may be written in a compact matrix form:

$$\boxed{\mathbf{V}=Z\mathbf{I}} \tag{22.4.4}$$

where Z is the *impedance matrix*. For example, in the case $K=4$, we have:

$$\mathbf{V}=\begin{bmatrix}V_1\\V_2\\V_3\\V_4\end{bmatrix}=\begin{bmatrix}Z_{11}&Z_{12}&Z_{13}&Z_{14}\\Z_{21}&Z_{22}&Z_{23}&Z_{24}\\Z_{31}&Z_{32}&Z_{33}&Z_{34}\\Z_{41}&Z_{42}&Z_{43}&Z_{44}\end{bmatrix}\begin{bmatrix}I_1\\I_2\\I_3\\I_4\end{bmatrix}=Z\mathbf{I}$$

We note that Z is a symmetric matrix, $Z=Z^T$, as a consequence of the reciprocity relations $Z_{pq}=Z_{qp}$.

Given the driving voltages V_p, Eq. (22.4.4) may be solved for the input currents I_p, which completely define the assumed sinusoidal currents $I_p(z)$ of Eq. (22.3.16). From the knowledge of the currents $I_p(z)$, one can obtain the radiation pattern of the array. Indeed, the radiation fields are obtained from Eq. (16.1.6), that is,

$$\boxed{\begin{aligned}\mathbf{E}&=\hat{\boldsymbol{\theta}}E_\theta=\hat{\boldsymbol{\theta}}\,jk\eta\,\frac{e^{-jkr}}{4\pi r}\,F_z(\theta,\phi)\sin\theta\\ \mathbf{H}&=\hat{\boldsymbol{\phi}}H_\phi=\hat{\boldsymbol{\phi}}\,jk\,\frac{e^{-jkr}}{4\pi r}\,F_z(\theta,\phi)\sin\theta\end{aligned}} \tag{22.4.5}$$

where the radiation vector $\mathbf{F}=\hat{\mathbf{z}}F_z$ has only a z-component given by:

$$F_z(\theta,\phi)=\int_{V'}J_z(\mathbf{r}')\,e^{j\mathbf{k}\cdot\mathbf{r}'}d\mathbf{r}' \tag{22.4.6}$$

But, in the thin-wire approximation, the total current density of the array is:

$$J_z(\mathbf{r}')=\sum_{p=1}^{K}I_p(z')\,\delta(x'-x_p)\,\delta(y'-y_p)$$

Inserting this into Eq. (22.4.6) and performing the x',y' integrations, we obtain:

$$F_z(\theta,\phi)=\sum_{p=1}^{K}e^{jk_xx_p+jk_yy_p}\int_{-h_p}^{h_p}I_p(z')\,e^{jk_zz'}dz' \tag{22.4.7}$$

Using Eq. (22.3.16) for $I_p(z)$ and replacing $k_z=k\cos\theta$, we obtain:

$$F_z(\theta,\phi)=\sum_{p=1}^{K}e^{jk_xx_p+jk_yy_p}\,\frac{2I_p}{k\sin kh_p}\,\frac{\cos(kh_p\cos\theta))-\cos kh_p}{\sin^2\theta} \tag{22.4.8}$$

The radiation intensity is given, in general, by Eq. (15.1.4):

$$U(\theta,\phi)=\frac{\eta k^2}{32\pi^2}\,\big|\sin\theta\,F_z(\theta,\phi)\big|^2$$

Replacing $k_x=k\sin\theta\cos\phi$ and $k_y=k\sin\theta\sin\phi$, we obtain:

$$\boxed{U(\theta,\phi)=\frac{\eta}{8\pi^2}\left|\sum_{p=1}^{K}I_p\,\frac{\cos(kh_p\cos\theta))-\cos kh_p}{\sin kh_p\sin\theta}\,e^{jk\sin\theta\,(x_p\cos\phi+y_p\sin\phi)}\right|^2}$$

Thus, the normalized *gain* of the array will be, up to a proportionality constant:

$$\boxed{g(\theta,\phi)=\left|\sum_{p=1}^{K}I_p\,\frac{\cos(kh_p\cos\theta))-\cos kh_p}{\sin kh_p\sin\theta}\,e^{jk\sin\theta\,(x_p\cos\phi+y_p\sin\phi)}\right|^2} \tag{22.4.9}$$

Equations (22.4.4) and (22.4.9) provide a complete solution to the problem of coupled antenna arrays, based on the sinusoidal approximation for the currents. In the special case of identical antennas, Eq. (22.4.9) factors as usual into an array factor and an element factor:

$$g(\theta,\phi)=\left|\sum_{p=1}^{K} I_p e^{jk\sin\theta\,(x_p\cos\phi+y_p\sin\phi)}\right|^2\left|\frac{\cos(kh\cos\theta))-\cos kh}{\sin kh\sin\theta}\right|^2$$

The MATLAB function `impedmat` calculates the $K{\times}K$ mutual impedance matrix $\mathcal{Z}$ of such an array, given the antenna lengths and radii, l_p, a_p, and the coordinates (x_p, y_p), for $p = 1, 2, \dots, K$. It has usage:

```
Z = impedmat(L,a,d);        % mutual impedance matrix of array of parallel dipoles
```

where all the lengths must be given in units of λ. It calls `imped` to calculate the individual matrix elements Z_{pq}.

The input parameters `L`, `a`, `d` are the vectors of antenna lengths, antenna radii, and (x_p, y_p) pairs, or the x_p positions, if the array is along the x-axis:

$$\boldsymbol{L}=\begin{bmatrix}L_1\\L_2\\\vdots\\L_K\end{bmatrix},\quad \mathbf{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_K\end{bmatrix},\quad \mathbf{d}=\begin{bmatrix}x_1,\ y_1\\x_2,\ y_2\\\vdots\\x_K,\ y_K\end{bmatrix}\quad\text{or}\quad\begin{bmatrix}x_1\\x_2\\\vdots\\x_K\end{bmatrix}$$

The MATLAB function `gain2s` calculates the E-plane and H-plane array gains using Eq. (22.4.9) and assumes that the input currents I_p have been obtained by solving Eq. (22.4.4). It has usage:

```
[ge,gh,th] = gain2s(L,d,I,N,ph0);    % gain of 2D array of dipoles with sinusoidal currents
[ge,gh,th] = gain2s(L,d,I,N);        % equivalent to φ0 = 0
```

where the input parameters `L`, `a` have the same meaning as in `impedmat`, and `I` is the vector of input currents $\boldsymbol{I} = [I_1, I_2, \dots, I_K]$. The output angle parameter `th` is either the polar or the azimuthal angle and takes N equally-spaced values in the interval $[0, 2\pi]$.

The H-plane gain $g_H(\phi)$ is defined to be the azimuthal gain on the xy-plane corresponding to $\theta = \pi/2$, and the E-plane gain $g_E(\theta)$ is defined to be the polar gain on any fixed azimuthal plane $\phi = \phi_0$, that is,

$$\boxed{\begin{aligned}g_H(\phi)&=g(\pi/2,\phi), && 0\le\phi\le 2\pi\\ g_E(\theta)&=g(\theta,\phi_0), && 0\le\theta\le 2\pi\end{aligned}}\tag{22.4.10}$$

Note that by allowing θ to vary over $[0, 2\pi]$, the E-plane gain can give both the forward and backward gain. The polar angle range $[0, \pi]$ covers the forward direction $\phi = \phi_0$, whereas, the range $[\pi, 2\pi]$ covers the backward direction $\phi = \phi_0 + \pi$, that is, we have the equivalence:

$$g(\theta,\phi_0)=g(\theta-\pi,\phi_0+\pi),\quad \pi\le\theta\le 2\pi$$

This follows from the trigonometric identities:

$$\begin{aligned}\sin(\theta-\pi)\cos(\phi_0+\pi)&=\sin\theta\cos\phi_0\\ \sin(\theta-\pi)\sin(\phi_0+\pi)&=\sin\theta\sin\phi_0\end{aligned}$$

Because both gains are defined over a 2π-angular range, they must be plotted with the MATLAB functions `abp2` and `abz2`, or in dB, with `dbp2` and `dbz2`.

Example 22.4.1: *Three-element parasitic array.* Undriven parasitic antennas located near transmitting ones can act as reflectors or directors, directing the radiation towards certain preferred directions. Fig. 22.4.2 shows an array of three half-wavelength dipoles. The geometry is the same as that of Example 19.3.3. The xy-coordinates of the elements are $\mathbf{d}_1 = (0, 0)$, $\mathbf{d}_2 = (0.5\lambda, 0)$, and $\mathbf{d}_3 = (0, 0.5\lambda)$.

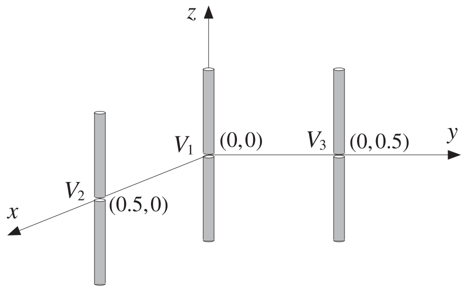


Fig. 22.4.2 Three-element array.

Let $\boldsymbol{V} = [V_1, V_2, V_3]^T$ be the driving voltages of the three elements. If only element-1 is driven and the others parasitic, we may take $\boldsymbol{V} = [1, 0, 0]^T$.

If the mutual couplings between the antennas are ignored, that is, the impedance matrix $\mathcal{Z}$ of Eq. (22.4.4) is taken to be diagonal, then, the input currents, will be $\boldsymbol{I} = [I_1, 0, 0]$ and the parasitic elements will be completely passive as though they were absent. The radiation pattern would be that of a single half-wave dipole. In particular, the azimuthal pattern would be omnidirectional.

This is not the case if the mutual couplings are taken into account. The parasitic elements act as reflectors, reflecting the radiation back towards the active element-1. By the symmetry of the arrangement, the maximum directivity will be in the direction with azimuthal angle $\phi = -135^{\text{o}}$. Fig. 22.4.3 shows the resulting H-plane and E-plane radiation patterns demonstrating this behavior. The dashed gains were computed by solving the coupled system of Hallén equations for the exact currents on each of the three antennas, as discussed in Example 22.6.1.

Assuming equal radii, $a = 0.001\lambda$, the 3×3 impedance matrix $\mathcal{Z}$ is found to be:

$$\mathcal{Z}=\begin{bmatrix}73.08+42.21j & -12.52-29.91j & -12.52-29.91j\\ -12.52-29.91j & 73.08+42.21j & -24.62+\ \ 0.78j\\ -12.52-29.91j & -24.62+\ \ 0.78j & 73.08+42.21j\end{bmatrix}$$

Then, the solution of Eq. (22.4.4) is:

$$\boldsymbol{I}=\begin{bmatrix}I_1\\I_2\\I_3\end{bmatrix}=\mathcal{Z}^{-1}\boldsymbol{V}=\mathcal{Z}^{-1}\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}0.0133\angle{-7.46^{\text{o}}}\\0.0066\angle 18.23^{\text{o}}\\0.0066\angle 18.23^{\text{o}}\end{bmatrix}$$

The typical MATLAB code used to generate these graphs was as follows:

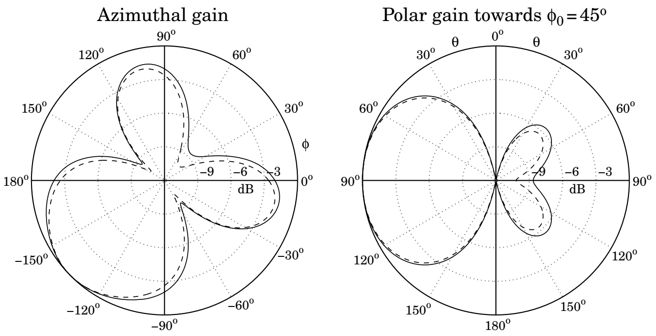


Fig. 22.4.3 *H*-plane and *E*-plane radiation patterns, $\boldsymbol{V} = [1, 0, 0]^T$.

```
L = [0.5, 0.5, 0.5];                      % lengths
a = [0.001, 0.001, 0.001];                % radii
d = [0,0; 0.5,0; 0,0.5];                  % xy locations

Z = impedmat(L,a,d);                      % impedance matrix
V = [1; 0; 0];                            % driving voltages
I = Z\V;                                  % input currents

ph0 = 45;                                 % 45° azimuthal plane for polar gain

[ge1,gh1,ph] = gain2s(L,d,I,360,ph0);     % gain2s assumes sinusoidal currents

M = 40;                                   % number of upper-half samples

[I,z] = hcoupled2(L,a,d,V,M);                 % solves for currents on all antennas

[ge2,gh2,ph] = gain2d(L,d,I,360,ph0);     % gain2d uses Hallén currents

figure; dbz2(ph,gh1,30,12); dbadd2(2,'--',ph,gh2,30,12);
figure; dbp2(ph,ge1,30,12); dbadd2(1,'--',ph,ge2,30,12);
```

Anticipating the symmetry about the 45° azimuthal plane, the *E*-plane gain was computed with $\phi_0 = 45^o$. As expected, the polar plot shows that the maximum gain is in the backward ϕ_0 direction, that is, toward $\phi_0 + 180^o = 225^o = -135^o$. □

Example 22.4.2: Next, consider the case when element-one is parasitic, but elements two and three are driven by equal voltages, $\boldsymbol{V} = [0, 1, 1]^T$. If the mutual coupling is ignored, then the two active elements act as an array which is broadside to the line joining them, that is, maximum directivity is in the 45° azimuthal direction, but with both the forward and the backward (i.e., −135°) directions being equal. This pattern is shown in the upper-right graph of Fig. 19.3.4.

If the mutual couplings are taken into account, element-1 will act as a reflector, reflecting towards the $\phi_0 = 45^o$ direction and reducing the gain in the opposite direction. This is demonstrated in Fig. 22.4.4. As in the previous example, the dashed gains correspond to the exact coupled Hallén solution.

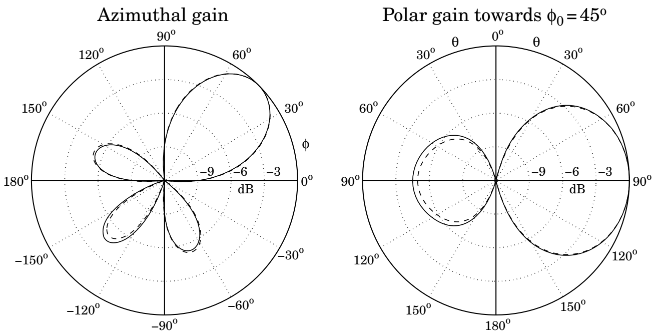


Fig. 22.4.4 *H*-plane and *E*-plane radiation patterns, $\boldsymbol{V} = [0, 1, 1]^T$.

Because of the identical geometry, the impedance matrix $\mathcal{Z}$ is the same as that of the previous example. But, the input currents are different:

$$\boldsymbol{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \mathcal{Z}^{-1}\boldsymbol{V} = \mathcal{Z}^{-1}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.0133\angle 18.23^o \\ 0.0173\angle{-19.04^o} \\ 0.0173\angle{-19.04^o} \end{bmatrix}$$

The only change in the previous MATLAB code was to use $\boldsymbol{V} = [0, 1, 1]^T$. □

Example 22.4.3: One of the earliest experimental studies of parasitic reflectors was by Nagy [1281]. One of his arrangements is shown in Fig. 22.4.5 in which the driven element is at the origin and the other three elements are parasitic. The antenna lengths were $l = 1.19$ m, and their radii $a = 0.395$ cm. The operating wavelength was $\lambda = 2.5$ meters, (i.e., frequency of 120 MHz.)

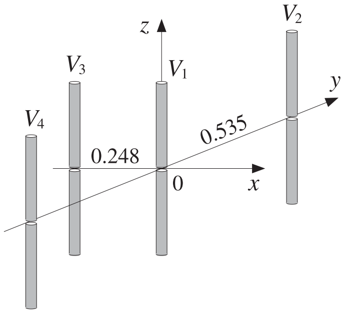


Fig. 22.4.5 Four-element parasitic array.

It follows that, $l = 0.476\lambda$ and $a = 0.00158\lambda$. Elements two and four were placed symmetrically along the y-axis at distances $\pm 0.535\lambda$, and element three was on the negative side of the x-axis at distance 0.248λ from the origin. Fig. 22.4.6 shows the calculated patterns. We observe that the three parasitic antennas act as reflectors, enhancing the radiation in the $\phi = 0$ direction.

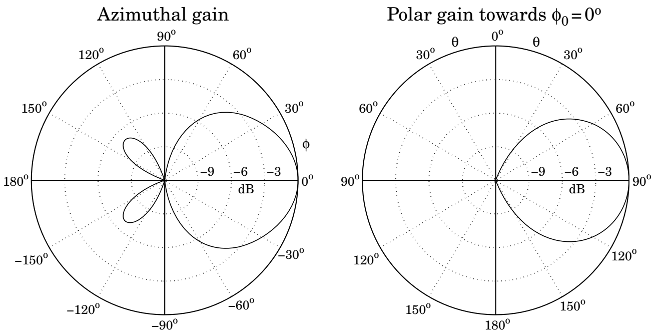


Fig. 22.4.6 H-plane and E-plane radiation patterns, $\mathbf{V} = [1, 0, 0, 0]^T$.

This array was later studied theoretically by Brown [1282], using the same methods as those presented here. Brown treated monopole antennas (i.e, half dipoles above a ground plane,) and therefore, the values of his mutual impedances are half of ours. The inputs to the design equations were the parameters:

$$\mathbf{L} = \begin{bmatrix} 0.476 \\ 0.476 \\ 0.476 \\ 0.476 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 0.00158 \\ 0.00158 \\ 0.00158 \\ 0.00158 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 0.000, & 0.000 \\ 0.000, & 0.535 \\ -0.248, & 0.000 \\ 0.000, & -0.535 \end{bmatrix}$$

The impedance matrix elements are:

$$\begin{aligned} &Z_{11} = Z_{22} = Z_{33} = 63.42\angle 0.65^\circ, \quad && Z_{12} = Z_{14} = 26.76\angle{-123.87^\circ} \\ &Z_{13} = 43.56\angle{-34.69^\circ}, && Z_{23} = Z_{34} = 24.78\angle{-141.96^\circ} \\ &Z_{24} = 14.74\angle 53.15^\circ && \end{aligned}$$

With $\mathbf{V} = [1, 0, 0, 0]^T$, the solution of $\mathcal{Z}\mathbf{I} = \mathbf{V}$ is:

$$\mathbf{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 0.0135\angle{-26.26^\circ} \\ 0.0043\angle 74.61^\circ \\ 0.0126\angle 116.70^\circ \\ 0.0043\angle 4.61^\circ \end{bmatrix}$$

and we find for the ratios:

$$\frac{I_2}{I_1} = \frac{I_4}{I_1} = 0.3180\angle 100.87^\circ, \quad \frac{I_3}{I_1} = 0.9343\angle 142.96^\circ$$

These numerical results are in close agreement with Brown's [1282]. The dashed Hallén gains are not shown, as in the previous examples, because they are virtually indistinguishable from the sinusoidal ones (for $M = 40$.) □

Example 22.4.4: *Coupled Dolph-Chebyshev array.* In this example, we study the impact of mutual coupling on the array design methods of Chap. 20. For a typical array spacing of half-wavelength, the mutual impedance matrix is *diagonally dominant* and therefore, there will be some but minor impact on the design.

Fig. 22.4.7 shows a 15-element array of z-directed half-wavelength dipoles with spacing $d = \lambda/2$ arranged along the x-axis. The antenna radii are $a = 0.001\lambda$.

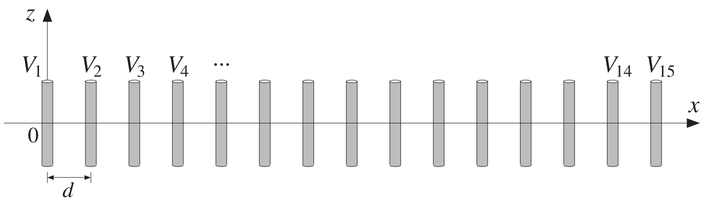


Fig. 22.4.7 Fifteen-element Dolph-Chebyshev array.

We take the feed voltages $\mathbf{V} = [V_1, V_2, \dots, V_{15}]^T$ to be Dolph-Chebyshev weights that would steer the azimuthal array gain towards $\phi_0 = 120^\circ$ and would achieve a 20-dB sidelobe level. These weights can be designed with the function `dolph`.

If the mutual coupling is ignored, the impedance matrix $\mathcal{Z}$ will be proportional to the identity matrix because the antenna elements are identical. Then, the input currents $\mathbf{I}$ will be essentially equal to the driving voltages $\mathbf{V}$ and the array will behave according to the desired design.

If the mutual coupling is taken into account, the currents must be calculated from the solution of $\mathcal{Z}\mathbf{I} = \mathbf{V}$ and some distortions on the desired angular pattern may occur because $\mathcal{Z}$ is no longer diagonal.

Fig. 22.4.8 shows the azimuthal and polar gain patterns with and without mutual coupling. The primary effect is to distort the sidelobe levels so that they are no longer equal. But they are still acceptable as a close approximation to the desired Dolph-Chebyshev pattern.

The typical MATLAB code used in this example was as follows:

```
K = 15;
ph0 = 120;                                  % steering angle

L = 0.5 * ones(1,K);                        % vector of antenna lengths
a = 0.001 * ones(1,K);                      % antenna radii
d = (0:K-1)*0.5;                            % equally-spaced with λ/2 spacing

V = dolph(0.5, ph0, K, 20).';               % Dolph design with 20-dB sidelobes

Z = impedmat(L,a,d);                        % 15×15 impedance matrix
I = Z\V;                                    % input currents

[ge,gh,ph] = gain2s(L,d,I,400,ph0);           % gains with coupling

figure; dbz2(ph,gh);                        % azimuthal gain
figure; dbp2(ph,ge);                        % polar gain
```

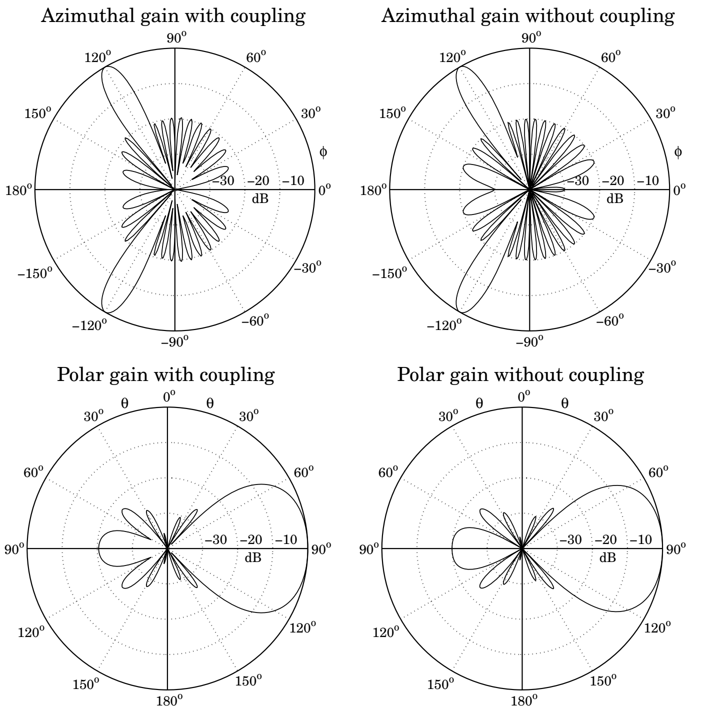


Fig. 22.4.8 *H*-plane and *E*-plane patterns with and without coupling.

```
[ge,gh,ph] = gain2s(L,d,V,400,ph0);                % gains without coupling

figure; dbz2(ph,gh);
figure; dbp2(ph,ge);
```

The *E*-plane polar gains were computed on the plane of the desired steering angle, that is, $\phi_0 = 120^{\rm o}$. The figures show that maximum gain is at $\theta = 90^{\rm o}$ in the ϕ_0 direction. In the case without coupling, we set $\boldsymbol{I} = \boldsymbol{V}$ inside `gain2s` because any proportionality constant gets canceled out. □

22.5 Yagi-Uda Antennas

A special type of parasitic array is the Yagi-Uda array shown in Fig. 22.5.1. The *z*-directed dipoles are arranged along the *x*-axis. The second dipole is driven; all others are parasitic.

The first dipole has length slightly longer than that of the driven dipole, and acts as a "reflector". The elements to the right of the driven dipole have lengths slightly shorter, and act as "directors." The reflector and directors direct the radiation preferentially towards endfire, that is, along the *x*-axis.

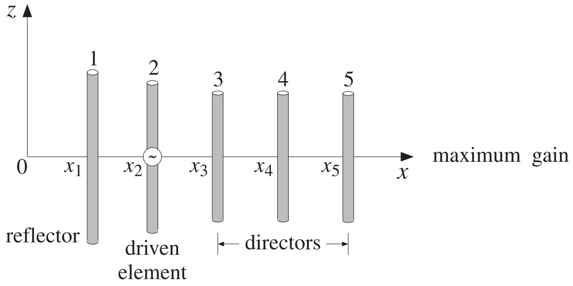


Fig. 22.5.1 Five-element Yagi-Uda array.

The Yagi-Uda array is widely used as a TV reception antenna and achieves fairly good directivity with such a simple structure. Good directivity characteristics are realized with certain choices for the antenna lengths and separations.

The analysis of the Yagi-Uda array follows the steps of the previous section. We assume that there are K dipoles, with the last $K-2$ being the directors, and that the currents are sinusoidal as in Eq. (22.3.16) because the antenna lengths are of the order of half-wavelength. Then, we compute the mutual impedance matrix $\mathcal{Z}$ and the input currents $\boldsymbol{I} = \mathcal{Z}^{-1}\boldsymbol{V}$. Because only the second element is driven, the vector of voltages is:

$$\boldsymbol{V} = [0, 1, \underbrace{0, 0, \dots, 0}_{(K-2)\ \text{zeros}}]^T \tag{22.5.1}$$

Once we have the input currents $\boldsymbol{I} = [I_1, I_2, \dots, I_K]^T$, the gain of the array is computed by Eq. (22.4.9), which simplifies into the following form because the dipoles lie along the *x*-axis:

$$\boxed{g(\theta,\phi) = \left| \sum_{p=1}^{K} I_p \frac{\cos(kh_p\cos\theta) - \cos kh_p}{\sin kh_p \sin\theta}\, e^{jkx_p \sin\theta\cos\phi} \right|^2} \tag{22.5.2}$$

We assume that the lengths and separations are such that the maximum gain is towards endfire, that is, towards $\theta = 90^{\rm o}$, $\phi = 0^{\rm o}$. The forward and backward gains, and the *forward-backward* or *front-to-back* ratio are defined as:

$$g_f = g_{\max} = g(90^{\rm o}, 0^{\rm o}), \quad g_b = g(90^{\rm o}, 180^{\rm o}), \quad R_{fb} = \frac{g_f}{g_b} \tag{22.5.3}$$

It follows that the normalized gain will be $g_n(\theta,\phi) = g(\theta,\phi)/g_f$. Integrating it over all solid angles, we obtain the beam solid angle and hence the directivity of the Yagi-Uda array:

$$\Delta\Omega = \int_0^{\pi}\int_0^{2\pi} g_n(\theta,\phi)\sin\theta\, d\theta\, d\phi, \qquad D = \frac{4\pi}{\Delta\Omega} \tag{22.5.4}$$

In dB, the directivity and forward-backward ratio are $10\log_{10}D$ and $10\log_{10}R_{fb}$. The MATLAB function `yagi` implements the above design steps. It computes the input currents $\boldsymbol{I}$ as well as the directivity and forward-backward ratio. Its usage is:

```
[I,D,Rfb] = yagi(L,a,d);                % Yagi-Uda array design
```

The function always assumes that the second element is the driven element and sets the value of $\boldsymbol{V}$ according to Eq. (22.5.1). The double integral in Eq. (22.5.4) is done with a 16-point Gauss-Legendre quadrature integration formula for each integration variable.

Example 22.5.1: *Reflectors and directors.* The simplest possible Yagi-Uda array has one driven element and either one reflector and no directors, or a single director and no reflector. Fig. 22.5.2 depicts the two cases.

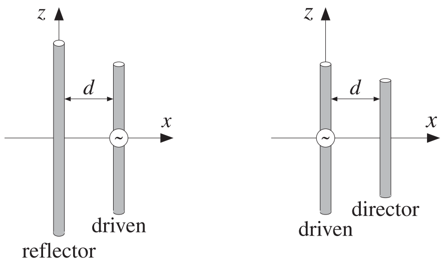


Fig. 22.5.2 The simplest Yagi-Uda arrays.

If the reflector is slightly longer than the driven element, and if the director is slightly shorter, then in both cases the radiation will be directed to the right, along the x-axis. Fig. 22.5.3 shows the resulting radiation patterns.

The length of the driven element was 0.50λ and that of the reflector and director, 0.54λ and 0.46λ, respectively. The antenna radii were $a = 0.003\lambda$ and their separation $d = 0.1\lambda$. The mutual impedances were calculated with `impedmat`:

$$Z = \begin{bmatrix} 92.47 + 104.19j & 75.68 + 11.63j \\ 75.68 + 11.63j & 73.07 + 41.37j \end{bmatrix}, \quad Z = \begin{bmatrix} 73.07 + 41.37j & 59.77 + 4.35j \\ 59.77 + 4.35j & 57.65 - 17.01j \end{bmatrix}$$

The typical MATLAB code that was used was:

```
L = [0.54,0.50]                          % reflector case
a = 0.003*[1,1];                         % radii
d = [0,0.1];                             % x-coordinates of locations

Z = impedmat(L,a,d);                     % impedance matrix
I = Z\[0,1]';                            % input currents

[ge,gh,th] = gain2s(L,d,I,400);          % gain computation

figure; dbz2(th,gh,30,16);               % azimuthal gain
figure; dbp2(th,ge,30,16);               % polar gain
```

The driving voltages were in the two cases: $\boldsymbol{V} = [0,1]^T$ and $\boldsymbol{V} = [1,0]^T$. □

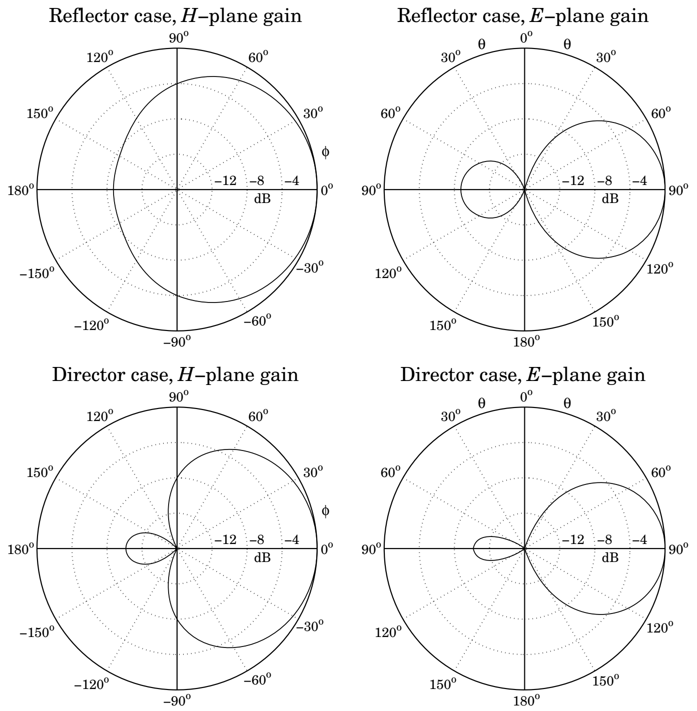

Reflector case, H–plane gain
Reflector case, E–plane gain
Director case, H–plane gain
Director case, E–plane gain

Fig. 22.5.3 H-plane and E-plane gains of simple Yagi-Uda arrays.

Example 22.5.2: *Three-element Yagi.* Here, we consider a three-element Yagi-Uda array with one reflector, one driven element, and one director. The corresponding antenna lengths, radii, and locations along the x-axis (with the driven element at the origin) were in units of λ:

$$\boldsymbol{L} = \begin{bmatrix} 0.50 \\ 0.48 \\ 0.46 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 0.003 \\ 0.003 \\ 0.003 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0 \\ 0.125 \end{bmatrix}$$

The azimuthal and polar gains are shown in Fig. 22.5.4. The dashed gains correspond to the exact coupled Hallén equations, as discussed in Example 22.6.3. The computed directivity and front/back ratio were $D = 8.18$ dB and $R_{fb} = 18.69$ dB. Thus, the array achieves a gain of $D - 2.15 = 6.03$ dB over a single half-wavelength dipole.

The impedance matrix was:

$$Z = \begin{bmatrix} 73.07 + 41.37j & 60.47 - 0.97j & 36.25 - 25.53j \\ 60.47 - 0.97j & 64.93 + 11.75j & 53.72 - 2.71j \\ 36.25 - 25.53j & 53.72 - 2.71j & 57.65 - 17.01j \end{bmatrix}$$

The input currents and input impedance of the driving element were:

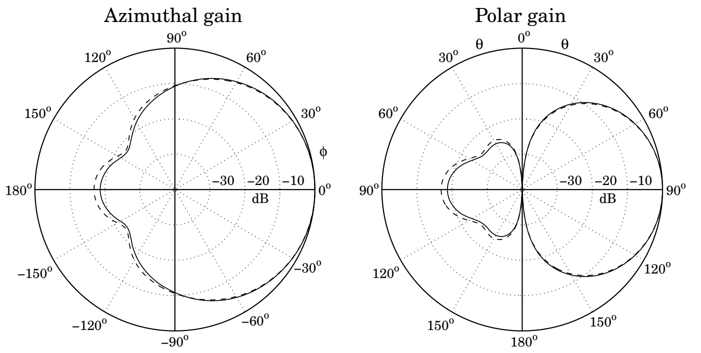


Fig. 22.5.4 Azimuthal and polar gains of three-element Yagi-Uda array.

$$\mathbf{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} -0.0290 + 0.0176j \\ 0.1062 - 0.0182j \\ -0.0801 - 0.0256j \end{bmatrix}, \quad Z_2 = \frac{V_2}{I_2} = \frac{1}{I_2} = 9.15 + 1.57j$$

The typical MATLAB code for this example was:

```
L = [0.50, 0.48, 0.46];                    % antenna lengths
a = 0.003*[1,1,1];                         % radii
d = [-0.125, 0, 0.125];                    % x-locations

[I,D,Rfb] = yagi(L,a,d);                   % solve ZI = V

[ge,gh,th] = gain2s(L,d,I,360);             % compute gains at 1° increments

M = 40;                                    % number of upper-half samples

[I,z] = hcoupled(L,a,d,[0,1,0],M);          % compute Hallén currents

[ge2,gh2,ph] = gain2d(L,d,I,360);          % gain of Hallén currents

figure; dbz2(ph,gh); dbadd2(2,'--',ph,gh2);
figure; dbp2(ph,ge); dbadd2(1,'--',ph,ge2);
```

The driving voltages were defined within `yagi` to be $\mathbf{V} = [0, 1, 0]^T$. □

Example 22.5.3: *Optimized six-element Yagi.* Chen and Cheng [1296] applied King's three-term current approximation [4] and devised procedures for optimizing the choices of the antenna lengths and separations of Yagi-Uda arrays. Fig. 22.5.5 shows the gains before and after optimization of a six-element Yagi-Uda array calculated with the functions `yagi` and `gain2s`. The antenna radii were $a = 0.003369\lambda$.

For the unoptimized case, the antenna lengths and x-locations were in units of λ:

$$\mathbf{L} = [L_1, L_2, L_3, L_4, L_5, L_6] = [0.510, 0.490, 0.430, 0.430, 0.430, 0.430]$$

$$\mathbf{d} = [x_1, x_2, x_3, x_4, x_5, x_6] = [-0.25, 0, 0.310, 0.620, 0.930, 1.240]$$

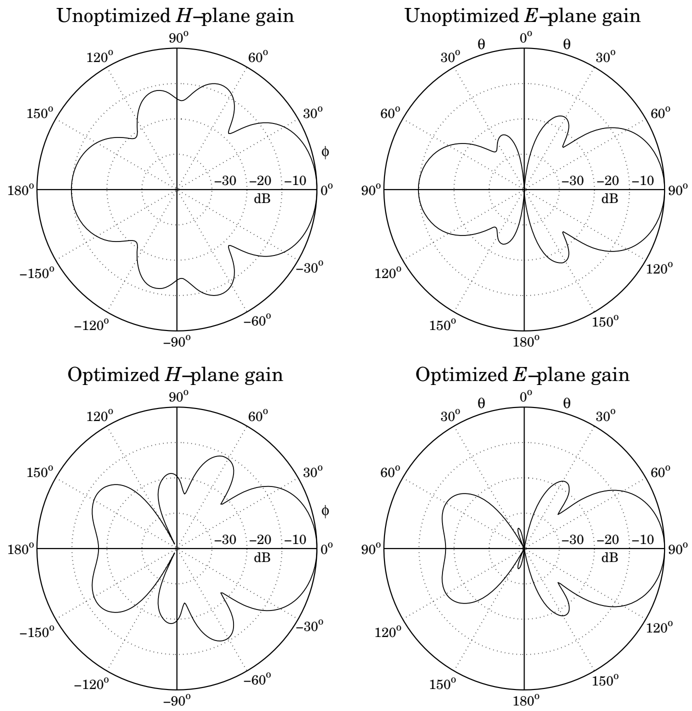


Fig. 22.5.5 Gains of six-element Yagi-Uda array.

The directors were identical and equally spaced at spacing of 0.31λ. The computed directivity and front/back ratio were 11 dB and 9.84 dB, respectively. The optimized case has slightly different lengths and x-locations:

$$\mathbf{L} = [L_1, L_2, L_3, L_4, L_5, L_6] = [0.476, 0.452, 0.436, 0.430, 0.434, 0.430]$$

$$\mathbf{d} = [x_1, x_2, x_3, x_4, x_5, x_6] = [-0.25, 0, 0.289, 0.695, 1.018, 1.440]$$

Typical MATLAB code was as follows:

```
L = [0.476, 0.452, 0.436, 0.430, 0.434, 0.430];
a = 0.003369 * [1,1,1,1,1,1];
d = [-0.25, 0, 0.289, 0.695, 1.018, 1.440];

[I,D,Rfb] = yagi(L,a,d);

[ge,gh,th] = gain2s(L,d,I,360);

figure; dbz2(th,gh,30,40);
figure; dbp2(th,ge,30,40);
```

The optimized directivity was 12.54 dB and the forward/backward ratio 17.6 dB. □

22.6 Hallén Equations for Coupled Antennas

In Sects. 22.3 and 22.4, we developed the Hallén-Pocklington equations for coupled antennas, that is, Eqs. (22.3.8)–(22.3.9) and (22.4.1). Here, we discuss their numerical solution. On the pth antenna, we have:

$$(\partial_z^2 + k^2)V_p(z) = 2kV_p\delta(z)\,, \quad p = 1, 2, \dots, K \tag{22.6.1}$$

where V_p is the driving delta-gap input and $V_p(z)$ is the sum of the (scaled) vector potentials due to the currents on all antennas:

$$\boxed{V_p(z) = \sum_{q=0}^{K} V_{pq}(z) = \frac{j\eta}{2\pi}\sum_{q=0}^{K}\int_{-h_q}^{h_q} G_{pq}(z-z')I_q(z')\,dz'} \tag{22.6.2}$$

where we recall the definition of the impedance kernel:

$$G_{pq}(z-z') = \frac{e^{-jkR}}{R}\,, \quad R = \sqrt{(z-z')^2 + d_{pq}^2} \tag{22.6.3}$$

and d_{pq} are the mutual distances or radii, as defined in Eq. (22.4.2). The use of the approximate kernel in (22.6.3) is well-justified for the off-diagonal terms $(p \neq q)$ because the distances d_{pq} are typically much greater than the radii. However, for the diagonal term $(p = q)$, one could use the exact kernel given as in (21.7.10) by,

$$G_{pp}(z-z') = \frac{2}{\pi}\int_0^{\pi/2}\frac{e^{-jkR}}{R}\,d\theta\,, \quad R = \sqrt{(z-z')^2 + 4a_p^2 - 4a_p^2\sin^2\theta} \tag{22.6.4}$$

Following the discussion of Sec. 21.3, the solution of (22.6.1) is of the form:

$$V_p(z) = C_p\cos kz + V_p\sin k|z|\,, \qquad -h_p \le z \le h_p \tag{22.6.5}$$

where we assumed that all the antennas are center-driven, and therefore, $V_p(z)$ will be even in z. Combining (22.6.5) with (22.6.2), we obtain the *coupled system of Hallén equations*, for $p = 1, 2, \dots, K$:

$$\boxed{\frac{j\eta}{2\pi}\sum_{q=0}^{K}\int_{-h_q}^{h_q} G_{pq}(z-z')I_q(z')\,dz' = C_p\cos kz + V_p\sin k|z|} \tag{22.6.6}$$

The K constants $C_1, C_2, \dots, C_K$ are determined by imposing the end conditions on the K currents: $I_p(h_p) = 0$, for $p = 1, 2, \dots, K$.

To solve this system, we use a basis-function expansion of the form of Eq. (21.8.3) and apply point matching. For simplicity, we take the same number of sampling points on each antenna, $N = 2M+1$. Because the antenna lengths may be different, the sample spacings will also be different. On the qth antenna we have,

$$\boxed{z_m = m\Delta_q\,, \quad \Delta_q = \frac{2h_q}{2M+1} = \frac{l_q}{N}}\,, \quad -M \le m \le M \tag{22.6.7}$$

The basis-function expansion for the qth current is

$$I_q(z') = \sum_{m=-M}^{M} I_q(z_m)B_q(z'-z_m)\,, \quad q = 1, 2, \dots, K \tag{22.6.8}$$

We are going to use only the pulse and the triangular bases defined by Eqs. (21.8.4) and (21.11.2), respectively, with sample spacing $\Delta = \Delta_q$. For the triangular basis, we must set $\Delta_q = h_q/M$ instead of $\Delta_q = 2h_q/(2M+1)$.

Inserting (22.6.8) into (22.6.6) and sampling along the p-th antenna, that is, at the points $z_n = n\Delta_p$, for $-M \le n \le M$, we obtain the discretized system:

$$\frac{j\eta}{2\pi}\sum_{q=0}^{K}\sum_{m=-M}^{M} I_q(z_m)\int_{-h_q}^{h_q} G_{pq}(z_n-z')B_q(z'-z_m)\,dz' = C_p\cos kz_n + V_p\sin k|z_n|$$

We define the $N\times N$ impedance matrix Z_{pq} whose nmth matrix element is:

$$\begin{aligned} Z_{pq}(n,m) &= \frac{j\eta}{2\pi}\int_{-h_q}^{h_q} G_{pq}(z_n-z')B_q(z'-z_m)\,dz' \\ &= \frac{j\eta}{2\pi}\int_{-h_q}^{h_q} G_{pq}(z_n-z_m-z)B_q(z)\,dz \end{aligned} \tag{22.6.9}$$

with $-M \le n, m \le M$. In particular, for the pulse basis, these take the form:

$$Z_{pq}(n,m) = \frac{j\eta}{2\pi}\int_{-\Delta_q/2}^{\Delta_q/2} G_{pq}(z_n-z_m-z)\,dz \tag{22.6.10}$$

and, for the triangular basis:

$$Z_{pq}(n,m) = \frac{j\eta}{2\pi}\int_{-\Delta_q}^{\Delta_q}\left(1 - \frac{|z|}{\Delta_q}\right)G_{pq}(z_n-z_m-z)\,dz \tag{22.6.11}$$

Denoting $I_q(m) = I_q(z_m)$, the discretized Hallén system becomes:

$$\boxed{\sum_{q=0}^{K}\sum_{m=-M}^{M} Z_{pq}(n,m)I_q(m) = C_p\cos kz_n + V_p\sin k|z_n|} \tag{22.6.12}$$

where $p = 1, 2, \dots, K$. And, written in a more compact form:

$$\boxed{\sum_{q=0}^{K} Z_{pq}\boldsymbol{I}_q = C_p\mathbf{c}_p + V_p\mathbf{s}_p} \tag{22.6.13}$$

where we defined the N-dimensional vectors:

$$\boldsymbol{I}_q = \begin{bmatrix} I_q(M) \\ \vdots \\ I_q(1) \\ I_q(0) \\ I_q(1) \\ \vdots \\ I_q(M) \end{bmatrix}, \quad \mathbf{c}_p = \begin{bmatrix} \cos kz_M \\ \vdots \\ \cos kz_1 \\ \cos kz_0 \\ \cos kz_1 \\ \vdots \\ \cos kz_M \end{bmatrix}, \quad \mathbf{s}_p = \begin{bmatrix} \sin kz_M \\ \vdots \\ \sin kz_1 \\ \sin kz_0 \\ \sin kz_1 \\ \vdots \\ \sin kz_M \end{bmatrix} \tag{22.6.14}$$

and used the even symmetry in z. The vectors $\mathbf{c}_p$ and $\mathbf{s}_p$ depend on p through the sample spacing in $z_n = n\Delta_p$, $-M \le n \le M$.

The system (22.6.13) provides K coupled matrix equations by which to determine the K sampled current vectors $\boldsymbol{I}_1, \boldsymbol{I}_2, \dots, \boldsymbol{I}_K$ on each antenna. The $N\times N$ matrices $\mathcal{Z}_{pq}$ are not Toeplitz, unless the antennas are identical, in which case $\Delta_p = \Delta_q$ and the $\mathcal{Z}_{pq}(n,m)$ depends only on the difference $n-m$. Of course, for $p=q$, $\mathcal{Z}_{pp}$ is both symmetric and Toeplitz.

However, while not Toeplitz, the matrix $\mathcal{Z}_{pq}$ is reversal-invariant because of the property $\mathcal{Z}_{pq}(n,m) = \mathcal{Z}_{pq}(-n,-m)$, which follows from Eq. (22.6.9). Therefore, the matrix system (22.6.13) can be wrapped in half by the procedure discussed in Sec. 21.9, which replaced the matrix equation (21.9.6) by (21.9.7).

Here, each $N\times N$ matrix $\mathcal{Z}_{pq}$ is wrapped to size $(M+1)\times(M+1)$ by the same process. The resulting system looks identical to (22.6.13), except the currents and right-hand sides are essentially half those of (22.6.14):

$$\boldsymbol{I}_q = \begin{bmatrix} I_q(0) \\ I_q(1) \\ \vdots \\ I_q(M) \end{bmatrix}, \quad \mathbf{c}_p = \begin{bmatrix} \cos kz_0 \\ \cos kz_1 \\ \vdots \\ \cos kz_M \end{bmatrix}, \quad \mathbf{s}_p = \begin{bmatrix} \sin kz_0 \\ \sin kz_1 \\ \vdots \\ \sin kz_M \end{bmatrix} \tag{22.6.15}$$

In particular, if all antennas are identical, then the wrapping process can be made even more efficient using the Toeplitz-Hankel properties of the wrapped matrices, as discussed in Example 21.9.1. In any case, we will assume in the sequel that the system (22.6.13) has been wrapped in half.

If the constants C_p were known, the solution of the system (22.6.13) could be obtained by writing it as a *single block-matrix* linear system of the form:

$$\boxed{\mathcal{Z}\boldsymbol{I} = C\mathbf{c} + V\mathbf{s}} \tag{22.6.16}$$

where $\mathcal{Z}$ is the $K\times K$ block matrix whose pqth matrix element is the $(M+1)\times(M+1)$ matrix $\mathcal{Z}_{pq}$, and C, V are appropriate block-diagonal matrices. The vectors $\boldsymbol{I}$, $\mathbf{c}$, $\mathbf{s}$ are the concatenations of $\boldsymbol{I}_p$, $\mathbf{c}_p$, $\mathbf{s}_p$. For example, in the case $K=3$, the system (22.6.13) reads:

$$\begin{aligned}
\mathcal{Z}_{11}\boldsymbol{I}_1 + \mathcal{Z}_{12}\boldsymbol{I}_2 + \mathcal{Z}_{13}\boldsymbol{I}_3 &= C_1\mathbf{c}_1 + V_1\mathbf{s}_1 \\
\mathcal{Z}_{21}\boldsymbol{I}_1 + \mathcal{Z}_{22}\boldsymbol{I}_2 + \mathcal{Z}_{23}\boldsymbol{I}_3 &= C_2\mathbf{c}_2 + V_2\mathbf{s}_2 \\
\mathcal{Z}_{31}\boldsymbol{I}_1 + \mathcal{Z}_{32}\boldsymbol{I}_2 + \mathcal{Z}_{33}\boldsymbol{I}_3 &= C_3\mathbf{c}_3 + V_3\mathbf{s}_3
\end{aligned}$$

This can be written in the 3×3 block-matrix form:

$$\begin{bmatrix} \mathcal{Z}_{11} & \mathcal{Z}_{12} & \mathcal{Z}_{13} \\ \mathcal{Z}_{21} & \mathcal{Z}_{22} & \mathcal{Z}_{23} \\ \mathcal{Z}_{31} & \mathcal{Z}_{32} & \mathcal{Z}_{33} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_1 \\ \boldsymbol{I}_2 \\ \boldsymbol{I}_3 \end{bmatrix} = \begin{bmatrix} C_1 I & 0 & 0 \\ 0 & C_2 I & \\ 0 & 0 & C_3 I \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix} + \begin{bmatrix} V_1 I & 0 & 0 \\ 0 & V_2 I & 0 \\ 0 & 0 & V_3 I \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \end{bmatrix}$$

where I is the $(M+1)\times(M+1)$ identity matrix.

Next, we discuss the determination of the constants C_p. The condition $I_p(M) = 0$ can be written vectorially in the form $\mathbf{u}^T\boldsymbol{I}_p = 0$, where $\mathbf{u} = [0,\dots,0,1]^T$, as was done in Sec. 21.9. Separating the pth term of the pth equation in (22.6.13), we have:

$$\mathcal{Z}_{pp}\boldsymbol{I}_p + \sum_{q\ne p}\mathcal{Z}_{pq}\boldsymbol{I}_q = C_p\mathbf{c}_p + V_p\mathbf{s}_p \tag{22.6.17}$$

Solving for $\boldsymbol{I}_p$ and multiplying by $\mathbf{u}^T$, we obtain the condition:

$$\mathbf{u}^T\boldsymbol{I}_p = C_p\mathbf{u}^T\mathcal{Z}_{pp}^{-1}\mathbf{c}_p + V_p\mathbf{u}^T\mathcal{Z}_{pp}^{-1}\mathbf{s}_p - \sum_{q\ne p}\mathbf{u}^T\mathcal{Z}_{pp}^{-1}\mathcal{Z}_{pq}\boldsymbol{I}_q = 0$$

Defining the quantity $\mathbf{u}_p = \mathcal{Z}_{pp}^{-1}\mathbf{u}$, we solve this condition for C_p:

$$C_p = \frac{1}{\mathbf{u}_p^T\mathbf{c}_p}\left(\sum_{q\ne p}\mathbf{u}_p^T\mathcal{Z}_{pq}\boldsymbol{I}_q - V_p\mathbf{u}_p^T\mathbf{s}_p\right)$$

Inserting C_p into Eq. (22.6.17) and rearranging terms, we obtain:

$$\mathcal{Z}_{pp}\boldsymbol{I}_p + \sum_{q\ne p}\left(I - \frac{\mathbf{c}_p\mathbf{u}_p^T}{\mathbf{c}_p^T\mathbf{u}_p}\right)\mathcal{Z}_{pq}\boldsymbol{I}_q = V_p\left(I - \frac{\mathbf{c}_p\mathbf{u}_p^T}{\mathbf{c}_p^T\mathbf{u}_p}\right)\mathbf{s}_p \tag{22.6.18}$$

To simplify it, we define the $(M+1)\times(M+1)$ *projection* matrices:

$$\mathcal{P}_p = I - \frac{\mathbf{c}_p\mathbf{u}_p^T}{\mathbf{c}_p^T\mathbf{u}_p}, \quad p = 1,2,\dots,K \tag{22.6.19}$$

Then, Eq. (22.6.18) can be written in the form:

$$\mathcal{Z}_{pp}\boldsymbol{I}_p + \sum_{q\ne p}\mathcal{P}_p\mathcal{Z}_{pq}\boldsymbol{I}_q = V_p\mathcal{P}_p\mathbf{s}_p \tag{22.6.20}$$

Thus, eliminating the constants C_p by enforcing the end conditions, amounts to replacing the impedance matrices $\mathcal{Z}_{pq}$ by the projected ones:

$$\bar{\mathcal{Z}}_{pq} = \begin{cases} \mathcal{Z}_{pp}, & \text{if } q = p \\ \mathcal{P}_p\mathcal{Z}_{pq}, & \text{if } q \ne p \end{cases} \tag{22.6.21}$$

and the term $\mathbf{s}_p$ by the projected one, $\bar{\mathbf{s}}_p = \mathcal{P}_p\mathbf{s}_p$. Then, Eq. (22.6.20) can be written in the form:

$$\boxed{\sum_{q=0}^{K}\bar{\mathcal{Z}}_{pq}\boldsymbol{I}_q = V_p\bar{\mathbf{s}}_p}, \quad p = 1,2,\dots,K \tag{22.6.22}$$

or, compactly in the block-matrix form:

$$\boxed{\bar{\mathcal{Z}}\boldsymbol{I} = V\bar{\mathbf{s}}} \tag{22.6.23}$$

with solution:

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{I}_1 \\ \boldsymbol{I}_2 \\ \vdots \\ \boldsymbol{I}_K \end{bmatrix} = \bar{\mathcal{Z}}^{-1}V\bar{\mathbf{s}} = \bar{\mathcal{Z}}^{-1}\begin{bmatrix} V_1\bar{\mathbf{s}}_1 \\ V_2\bar{\mathbf{s}}_2 \\ \vdots \\ V_K\bar{\mathbf{s}}_K \end{bmatrix} \tag{22.6.24}$$

The MATLAB function `hcoupled` implements the above solution procedure. First, it constructs the impedance matrices $\mathcal{Z}_{pq}$ by calculating the integrals in Eq. (22.6.9) using a 32-point Gauss-Legendre quadrature integration formula. Second, it wraps the matrices $\mathcal{Z}_{pq}$ in half and puts them together into the block-matrix $\mathcal{Z}$. And third, it constructs the projected matrix $\bar{\mathcal{Z}}$ and the solution (22.6.24). Its usage is:

```
[I,z] = hcoupled(L,a,d,V,M,ker,basis);     % solve Hallén equations for coupled dipoles
```

where $\boldsymbol{L}, \mathbf{a}, \mathbf{d}$ are the vectors of antenna lengths, radii, and xy-locations, and $\boldsymbol{V}$ is the vector of the driving voltages $\boldsymbol{V} = [V_1, V_2, \dots, V_K]$. The parameters $\boldsymbol{L}, \mathbf{a}, \mathbf{d}$ have the same usage as in the functions `yagi` and `gain2s`. The string input `ker` takes the values `'e','a'` for using the exact or the approximate kernel in the computation of the diagonal elements of the impedance matrix. The `basis` parameter can take only the two values `'p','t'` for pulse or triangular basis.

The output $\boldsymbol{I}$ is the $(2M+1)\times K$ matrix whose pth column is the double-sided vector of current samples $I_p(z_m)$, $z_m = m\Delta_p$, $-M \le m \le M$. Thus, the matrix elements of $\boldsymbol{I}$ are $I(m,p) = I_p(z_m)$. Similarly, the pth column of the output matrix $\mathbf{z}$ holds the sampled z-locations on the pth antenna, that is, $z(m,p) = m\Delta_p$.

The output matrix $\boldsymbol{I}$ is obtained by using the MATLAB function `reshape` to reshape the $\big(K(M+1)\big)$-dimensional column vector solution (22.6.24) into a matrix of size $(M+1)\times K$, and then, symmetrizing it to size $(2M+1)\times K$.

A faster version of `hcoupled` is the function `hcoupled2`, which assumes that the antennas are identical. It is faster because it makes use of the Toeplitz-Hankel structure of the wrapped matrices $\mathcal{Z}_{pq}$ to construct them more efficiently. Its usage is:

```
[I,z] = hcoupled2(L,a,d,V,M,ker,basis));    % Hallén equations for coupled identical dipoles
```

where $\boldsymbol{I}$ has the same meaning as in `hcoupled`, but $\mathbf{z}$ is now a single column vector, that is, $z_m = m\Delta$, $-M \le m \le M$. In both `hcoupled2` and `hcoupled`, the final solution is obtained by solving the system (22.6.23), which is $\big(K(M+1)\big)\times\big(K(M+1)\big)$-dimensional.

In order to conveniently manipulate the block impedance matrices, we developed a MATLAB function, `blockmat`, which is used extensively inside `hcoupled2` and `hcoupled`. It allows one to create block matrices and to extract or insert sub-blocks. Its usage is as follows:

```
Z   = blockmat(K,K,M+1,M+1);      % create a (K(M+1))×(K(M+1)) matrix of zeros
Zpq = blockmat(K,K,p,q,Z);        % extract pqth submatrix of Z
Z   = blockmat(K,K,p,q,Z,Zpq);    % insert Zpq into pqth submatrix of Z
s   = blockmat(K,1,M+1,1);        % create a (K(M+1))-dimensional column of zeros
sp  = blockmat(K,1,p,1,s);        % extract the pth subvector of s
s   = blockmat(K,1,p,1,s,sp);     % insert sp into pth subvector of s
```

Once the sampled currents $I_p(m)$ are known, the gain of the array can be computed by finding the total current density, $\boldsymbol{J}(\mathbf{r}) = \hat{\mathbf{z}}J_z(\mathbf{r})$:

$$J_z(\mathbf{r}) = \sum_{p=1}^{K} I_p(z)\,\delta(x-x_p)\,\delta(y-y_p) = \sum_{p=1}^{K}\sum_{m=-M}^{M} I_p(m)\,B_p(z-z_m)\,\delta(x-x_p)\,\delta(y-y_p)$$

where we used Eq. (22.6.8). The corresponding radiation vector is:

$$\begin{aligned} F_z(\theta,\phi) &= \int J_z(\mathbf{r})\,e^{j\mathbf{k}\cdot\mathbf{r}}\,d^3\mathbf{r} = \sum_{m=-M}^{M}\sum_{p=1}^{K} I_p(m)\,e^{jk_x x_p + jk_y y_p}\int_{-h_p}^{h_p} B_p(z-z_m)\,e^{jk_z z}\,dz \\ &= \sum_{m=-M}^{M}\sum_{p=1}^{K} I_p(m)\,e^{jk_x x_p + jk_y y_p}\,e^{jk_z z_m}\int_{-h_p}^{h_p} B_p(z)\,e^{jk_z z}\,dz \end{aligned}$$

Performing the z-integration, we finally get:

$$\boxed{F_z(\theta,\phi) = \sum_{m=-M}^{M}\sum_{p=1}^{K} I_p(m)\,e^{jk_z m\Delta_p}\,e^{jk_x x_p + jk_y y_p}\,\Delta_p\left[\frac{\sin(k_z\Delta_p/2)}{k_z\Delta_p/2}\right]^b} \tag{22.6.25}$$

where $k_x = k\sin\theta\cos\phi$, $k_y = k\sin\theta\sin\phi$, and $k_z = k\cos\theta$, and $b = 1$ for the pulse basis and $b = 2$ for the triangular one. The corresponding normalized gain of the array will be, up to a constant:

$$g(\theta,\phi) = \big|\sin\theta\,F_z(\theta,\phi)\big|^2 \tag{22.6.26}$$

The MATLAB function `gain2d` computes the E-plane polar gain and the H-plane azimuthal gain from Eqs. (22.6.25) and (22.6.26). Its usage is:

```
[ge,gh,th] = gain2d(L,d,I,N,ph0,basis)    % gain of 2D array of antennas with Hallén currents
[ge,gh,th] = gain2d(L,d,I,N,ph0)          % equivalent to basis='p' (pulse basis)
[ge,gh,th] = gain2d(L,d,I,N,basis)        % equivalent to ph0=0
[ge,gh,th] = gain2d(L,d,I,N)              % equivalent to ph0=0, basis='p'
```

where the current input $\boldsymbol{I}$ is exactly the same as the output matrix from `hcoupled` or `hcoupled2`. The meaning of the outputs are exactly the same as in the function `gain2s` discussed in Sec. 22.4. The string `basis` takes the values `'p'` or `'t'`, for pulse or triangular basis.

The difference between `gain2s` and `gain2d` is that the former assumes the currents are sinusoidal and $\boldsymbol{I}$ represents only the input currents, $\boldsymbol{I} = [I_1, I_2, \dots, I_K]$. whereas in the latter, the full $(2M+1)\times K$ current matrix is needed, $\boldsymbol{I} = [\boldsymbol{I}_1, \boldsymbol{I}_2, \dots, \boldsymbol{I}_K]$.

Example 22.6.1: *Hallén solution of parasitic array.* Consider the three-element array of Example 22.4.1 and shown in Fig. 22.4.2. The Hallén currents on each antenna can be computed by using `hcoupled2` because the elements are identical. Fig. 22.6.1 shows the computed sampled currents with $N = 2M+1 = 81$ or $M = 40$.

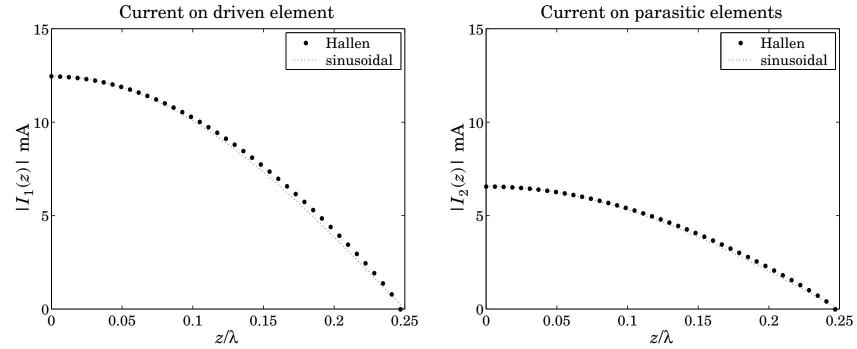


Fig. 22.6.1 Currents on driven and parasitic antennas.

Because of the symmetry, the currents on the two parasitic antennas are the same. For all three antennas, the currents are essentially sinusoidal, justifying the use of this assumption. The gains computed with `gain2d`, and under the sinusoidal assumption with `gain2s`, were shown in Fig. 22.4.3. The MATLAB code used to generate the currents and the gains was given in Example 22.4.1. □

Example 22.6.2: *Full-wavelength parasitic array.* If one or more of the antennas has length equal to a multiple of λ, the analysis methods based on the sinusoidal assumption break down because the impedance matrix computed with Eq. (22.3.15) becomes infinite.

On the other hand, the numerical solution of the Hallén system can still be carried through giving a finite answer. Fig. 22.6.2 shows the gains and currents of the parasitic array of Example 22.4.1, but all the antennas being full-wavelength elements, $l = \lambda$. The distance of the parasitic antennas to the driven element was also changed to $d = 0.25\lambda$ from $d = 0.5\lambda$.

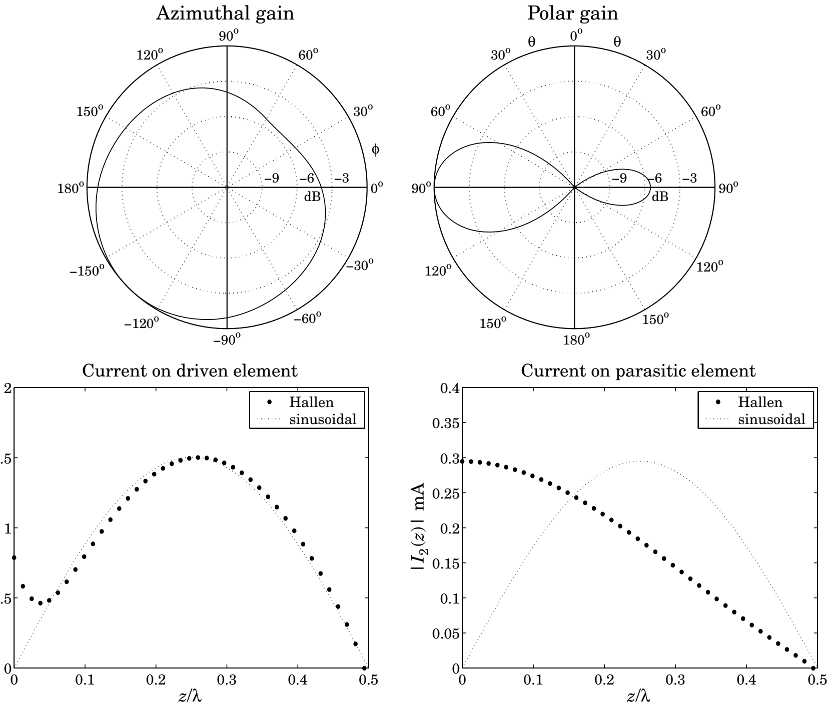


Fig. 22.6.2 Gains and currents of full-wavelength parasitic array.

The sinusoidal assumption for the driven element is fairly accurate except near $z = 0$, where the current has an non-zero value. But on the parasitic element, the sinusoidal assumption is completely wrong. □

Example 22.6.3: *Three-element Yagi-Uda array.* Here, we compute the currents on the three antennas of the Yagi-Uda array of Example 22.5.2. Because the antennas are not identical, the function `hcoupled` must be used. The gains were computed with `gain2s` and `gain2d` in Example 22.5.2. and shown in Fig. 22.5.4. The sampled currents on the three antennas are shown in Fig. 22.6.3.

We observe that the sinusoidal assumption is fairly accurate. The MATLAB code used to generate the current graphs was as follows:

```
L = [0.50, 0.48, 0.46]; h = L/2;
a = 0.003 * [1, 1, 1];
d = [-0.125, 0, 0.125];
V = [0, 1, 0];                          % can be defined as column or row

k = 2*pi;
M = 40;

[I,z] = hcoupled(L,a,d,V,M);

I1 = abs(I(M+1:end,1));  m1 = max(I1);  z1 = z(M+1:end,1);
I2 = abs(I(M+1:end,2));  m2 = max(I2);  z2 = z(M+1:end,2);
I3 = abs(I(M+1:end,3));  m3 = max(I3);  z3 = z(M+1:end,3);

s1 = 0:h(1)/50:h(1);  Is1 = m1*abs(sin(k*(h(1)-s1)));
s2 = 0:h(2)/50:h(2);  Is2 = m2*abs(sin(k*(h(2)-s2)));
s3 = 0:h(3)/50:h(3);  Is3 = m3*abs(sin(k*(h(3)-s3)));

figure;  plot(z1, I1, '.', s1, Is1, ':');
figure;  plot(z2, I2, '.', s2, Is2, ':');
figure;  plot(z3, I3, '.', s3, Is3, ':');
```

Note that $\mathbf{I}_1$, $\mathbf{I}_2$, and $\mathbf{I}_3$ are obtained from the three columns of $\mathbf{I}$, and $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ from the three columns of $\mathbf{z}$. Only the currents on the upper-half of each antenna are plotted. The sinusoidal currents are scaled to the maximum values of the corresponding Hallén currents.

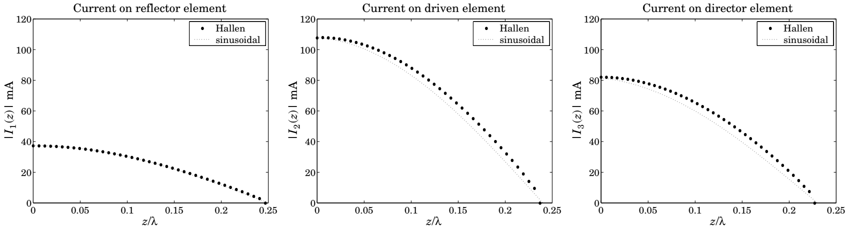


Fig. 22.6.3 Currents on the Yagi antennas.

These examples demonstrate the remark made earlier that the sinusoidal assumption is justified only for antennas with lengths near half a wavelength. □

22.7 Problems

22.1 Show that the asymptotic form of Eq. (22.2.7) for the mutual impedance between two parallel dipoles separated by a distance d is given by

$$Z_{21} = \frac{j\eta}{\pi}\tan\left(\frac{kh_1}{2}\right)\tan\left(\frac{kh_2}{2}\right)\frac{e^{-jkd}}{kd}\,, \quad \text{for large } d$$

22.2 Using the higher-order terms in the series (22.2.23), show that the input impedance $Z_{\text{in}} = R + jX$ of a small dipole is given as follows to order $(kl)^4$, where $L = \ln(2a/l)$:

$$R = \frac{\eta}{2\pi}\left[\frac{1}{12}(kl)^2 + \frac{1}{360}(kl)^4\right], \quad X = \frac{\eta}{2\pi}\left[\frac{4(1+L)}{kl} - \frac{1}{3}\left(L + \frac{2}{3}\right)(kl) - \frac{1}{180}\left(L - \frac{11}{30}\right)(kh)^3\right]$$

22.3 Consider a small dipole with a linear current given by Eq. (22.2.25). Determine the radiation vector, and the radiated electric and magnetic fields at a far distance r from the dipole. Calculate the radiated power P_{rad} by integrating the radial Poynting vector over a large sphere. Then identify the radiation resistance R through the definition:

$$P_{\text{rad}} = \frac{1}{2}R|I_0|^2$$

and show R is the same as that given by Eq. (22.2.24)

23

Appendices

A. Physical Constants

We use SI units throughout this text. Simple ways to convert between SI and other popular units, such as Gaussian, may be found in Refs. [100–103].

The Committee on Data for Science and Technology (CODATA) of NIST maintains the values of many physical constants [89]. The most current values can be obtained from the CODATA web site [1299]. Some commonly used constants are listed below:

quantity	**symbol**	**value**	**units**
speed of light in vacuum	c_0, c	299 792 458	$\mathrm{m\,s^{-1}}$
permittivity of vacuum	ϵ_0	$8.854\,187\,817 \times 10^{-12}$	$\mathrm{F\ m^{-1}}$
permeability of vacuum	μ_0	$4\pi \times 10^{-7}$	$\mathrm{H\ m^{-1}}$
characteristic impedance	η_0, Z_0	376.730 313 461	Ω
electron charge	e	$1.602\,176\,462 \times 10^{-19}$	C
electron mass	m_e	$9.109\,381\,887 \times 10^{-31}$	kg
Boltzmann constant	k	$1.380\,650\,324 \times 10^{-23}$	$\mathrm{J\,K^{-1}}$
Avogadro constant	N_A, L	$6.022\,141\,994 \times 10^{23}$	$\mathrm{mol^{-1}}$
Planck constant	h	$6.626\,068\,76 \times 10^{-34}$	J/Hz
Gravitational constant	G	$6.672\,59 \times 10^{-11}$	$\mathrm{m^3\,kg^{-1}\,s^{-2}}$
Earth mass	$M_\oplus$	5.972×10^{24}	kg
Earth equatorial radius	a_e	6378	km

In the table, the constants c, μ_0 are taken to be exact, whereas ϵ_0, η_0 are derived from the relationships:

$$\epsilon_0 = \frac{1}{\mu_0 c^2}\,, \quad \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = \mu_0 c$$

The energy unit of electron volt (eV) is defined to be the work done by an electron in moving across a voltage of one volt, that is, 1 eV = $1.602\,176\,462 \times 10^{-19}$ C · 1 V, or

$$1\ \mathrm{eV} = 1.602\,176\,462 \times 10^{-19}\ \mathrm{J}$$

In units of eV/Hz, Planck's constant h is:

$$h = 4.135\,667\,27 \times 10^{-15}\ \text{eV/Hz} = 1\ \text{eV}/241.8\ \text{THz}$$

that is, 1 eV corresponds to a frequency of 241.8 THz, or a wavelength of 1.24 μm.

B. Electromagnetic Frequency Bands

The ITU[†] divides the radio frequency (RF) spectrum into the following frequency and wavelength bands in the range from 30 Hz to 3000 GHz:

RF Spectrum

	band designations	**frequency**		**wavelength**	
ELF	Extremely Low Frequency	30–300	Hz	1–10	Mm
VF	Voice Frequency	300–3000	Hz	100–1000	km
VLF	Very Low Frequency	3–30	kHz	10–100	km
LF	Low Frequency	30–300	kHz	1–10	km
MF	Medium Frequency	300–3000	kHz	100–1000	m
HF	High Frequency	3–30	MHz	10–100	m
VHF	Very High Frequency	30–300	MHz	1–10	m
UHF	Ultra High Frequency	300–3000	MHz	10–100	cm
SHF	Super High Frequency	3–30	GHz	1–10	cm
EHF	Extremely High Frequency	30–300	GHz	1–10	mm
	Submillimeter	300-3000	GHz	100–1000	μm

An alternative subdivision of the low-frequency bands is to designate the bands 3–30 Hz, 30–300 Hz, and 300–3000 Hz as extremely low frequency (ELF), super low frequency (SLF), and ultra low frequency (ULF), respectively.

Microwaves span the 300 MHz–300 GHz frequency range. Typical microwave and satellite communication systems and radar use the 1–30 GHz band. The 30–300 GHz EHF band is also referred to as the millimeter band.

The 1–100 GHz range is subdivided further into the subbands shown on the right.

Microwave Bands

band	**frequency**	
L	1–2	GHz
S	2–4	GHz
C	4–8	GHz
X	8–12	GHz
Ku	12–18	GHz
K	18–27	GHz
Ka	27–40	GHz
V	40–75	GHz
W	80–100	GHz

Some typical RF applications are as follows. AM radio is broadcast at 535–1700 kHz falling within the MF band. The HF band is used in short-wave radio, navigation, amateur, and CB bands. FM radio at 88–108 MHz, ordinary TV, police, walkie-talkies, and remote control occupy the VHF band.

Cell phones, personal communication systems (PCS), pagers, cordless phones, global positioning systems (GPS), RF identification systems (RFID), UHF-TV channels, microwave ovens, and long-range surveillance radar fall within the UHF band.

[†] International Telecommunication Union.

The SHF microwave band is used in radar (traffic control, surveillance, tracking, missile guidance, mapping, weather), satellite communications, direct-broadcast satellite (DBS), and microwave relay systems. Multipoint multichannel (MMDS) and local multipoint (LMDS) distribution services, fall within UHF and SHF at 2.5 GHz and 30 GHz.

Industrial, scientific, and medical (ISM) bands are within the UHF and low SHF, at 900 MHz, 2.4 GHz, and 5.8 GHz. Radio astronomy occupies several bands, from UHF to L–W microwave bands.

Beyond RF, come the infrared (IR), visible, ultraviolet (UV), X-ray, and γ-ray bands. The IR range extends over 3–300 THz, or 1–100 μm. Many IR applications fall in the 1–20 μm band. For example, optical fiber communications typically use laser light at 1.55 μm or 193 THz because of the low fiber losses at that frequency. The UV range lies beyond the visible band, extending typically over 10–400 nm.

band	**wavelength**	**frequency**	**energy**
infrared	100–1 μm	3–300 THz	
ultraviolet	400–10 nm	750 THz–30 PHz	
X-Ray	10 nm–100 pm	30 PHz–3 EHz	0.124–124 keV
γ-ray	< 100 pm	> 3 EHz	> 124 keV

The CIE[†] defines the visible spectrum to be the wavelength range 380–780 nm, or 385–789 THz. Colors fall within the following typical wavelength/frequency ranges:

Visible Spectrum

color	**wavelength**	**frequency**
red	780–620 nm	385–484 THz
orange	620–600 nm	484–500 THz
yellow	600–580 nm	500–517 THz
green	580–490 nm	517–612 THz
blue	490–450 nm	612–667 THz
violet	450–380 nm	667–789 THz

X-ray frequencies fall in the PHz (petahertz) range and γ-ray frequencies in the EHz (exahertz) range.[‡] X-rays and γ-rays are best described in terms of their energy, which is related to frequency through Planck's relationship, $E = hf$. X-rays have typical energies of the order of keV, and γ-rays, of the order of MeV and beyond. By comparison, photons in the visible spectrum have energies of a couple of eV.

The earth's atmosphere is mostly opaque to electromagnetic radiation, except for three significant "windows", the visible, the infrared, and the radio windows. These three bands span the wavelength ranges of 380-780 nm, 1-12 μm, and 5 mm–20 m, respectively.

Within the 1-10 μm infrared band there are some narrow transparent windows. For the rest of the IR range (1–1000 μm), water and carbon dioxide molecules absorb infrared radiation—this is responsible for the Greenhouse effect. There are also some minor transparent windows for 17–40 and 330–370 μm.

[†] Commission Internationale de l'Eclairage (International Commission on Illumination.)

[‡] 1 THz = 10^{12} Hz, 1 PHz = 10^{15} Hz, 1 EHz = 10^{18} Hz.

Beyond the visible band, ultraviolet and X-ray radiation are absorbed by ozone and molecular oxygen (except for the ozone holes.)

C. Vector Identities and Integral Theorems

Algebraic Identities

$$|\boldsymbol{A}|^2|\boldsymbol{B}|^2 = |\boldsymbol{A}\cdot\boldsymbol{B}|^2 + |\boldsymbol{A}\times\boldsymbol{B}|^2 \tag{C.1}$$

$$(\boldsymbol{A}\times\boldsymbol{B})\cdot\boldsymbol{C} = (\boldsymbol{B}\times\boldsymbol{C})\cdot\boldsymbol{A} = (\boldsymbol{C}\times\boldsymbol{A})\cdot\boldsymbol{B} \tag{C.2}$$

$$\boldsymbol{A}\times(\boldsymbol{B}\times\boldsymbol{C}) = \boldsymbol{B}\,(\boldsymbol{A}\cdot\boldsymbol{C}) - \boldsymbol{C}\,(\boldsymbol{A}\cdot\boldsymbol{B}) \qquad \text{(BAC-CAB rule)} \tag{C.3}$$

$$(\boldsymbol{A}\times\boldsymbol{B})\cdot(\boldsymbol{C}\times\boldsymbol{D}) = (\boldsymbol{A}\cdot\boldsymbol{C})\,(\boldsymbol{B}\cdot\boldsymbol{D}) - (\boldsymbol{A}\cdot\boldsymbol{D})\,(\boldsymbol{B}\cdot\boldsymbol{C}) \tag{C.4}$$

$$(\boldsymbol{A}\times\boldsymbol{B})\times(\boldsymbol{C}\times\boldsymbol{D}) = \big[(\boldsymbol{A}\times\boldsymbol{B})\cdot\boldsymbol{D}\big]\boldsymbol{C} - \big[(\boldsymbol{A}\times\boldsymbol{B})\cdot\boldsymbol{C}\big]\boldsymbol{D} \tag{C.5}$$

$$\boldsymbol{A} = \hat{\mathbf{n}}\times(\boldsymbol{A}\times\hat{\mathbf{n}}) + (\hat{\mathbf{n}}\cdot\boldsymbol{A})\hat{\mathbf{n}} = \boldsymbol{A}_\perp + \boldsymbol{A}_\parallel \tag{C.6}$$

where $\hat{\mathbf{n}}$ is any unit vector, and $\boldsymbol{A}_\perp$, $\boldsymbol{A}_\parallel$ are the components of $\boldsymbol{A}$ perpendicular and parallel to $\hat{\mathbf{n}}$. Note also that $\hat{\mathbf{n}}\times(\boldsymbol{A}\times\hat{\mathbf{n}}) = (\hat{\mathbf{n}}\times\boldsymbol{A})\times\hat{\mathbf{n}}$. A three-dimensional vector can equally well be represented as a column vector:

$$\boldsymbol{a} = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}} \quad\Leftrightarrow\quad \boldsymbol{a} = \begin{bmatrix} a_x \\ a_y \\ b_z \end{bmatrix} \tag{C.7}$$

Consequently, the dot and cross products may be represented in matrix form:

$$\boldsymbol{a}\cdot\boldsymbol{b} \quad\Leftrightarrow\quad \boldsymbol{a}^T\boldsymbol{b} = [a_x, a_y, a_z]\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = a_xb_x + a_yb_y + a_zb_z \tag{C.8}$$

$$\boldsymbol{a}\times\boldsymbol{b} \quad\Leftrightarrow\quad A\boldsymbol{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \begin{bmatrix} a_yb_z - a_zb_y \\ a_zb_x - a_xb_z \\ a_xb_y - a_yb_x \end{bmatrix} \tag{C.9}$$

The cross-product matrix A satisfies the following identity:

$$A^2 = \boldsymbol{a}\boldsymbol{a}^T - (\boldsymbol{a}^T\boldsymbol{a})I \tag{C.10}$$

where I is the 3×3 identity matrix. Applied to a unit vector $\hat{\mathbf{n}}$, this identity reads:

$$I = \hat{\mathbf{n}}\hat{\mathbf{n}}^T - \hat{N}^2\,, \quad \text{where} \quad \hat{\mathbf{n}} = \begin{bmatrix} \hat{n}_x \\ \hat{n}_y \\ \hat{n}_z \end{bmatrix}, \quad \hat{N} = \begin{bmatrix} 0 & -\hat{n}_z & \hat{n}_y \\ \hat{n}_z & 0 & -\hat{n}_x \\ -\hat{n}_y & \hat{n}_x & 0 \end{bmatrix}, \quad \hat{\mathbf{n}}^T\hat{\mathbf{n}} = 1 \tag{C.11}$$

This corresponds to the matrix form of the parallel/transverse decomposition (C.6). Indeed, we have $\boldsymbol{a}_\parallel = \hat{\mathbf{n}}(\hat{\mathbf{n}}^T\boldsymbol{a})$ and $\boldsymbol{a}_\perp = (\hat{\mathbf{n}}\times\boldsymbol{a})\times\hat{\mathbf{n}} = -\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\boldsymbol{a}) = -\hat{N}(\hat{N}\boldsymbol{a}) = -\hat{N}^2\boldsymbol{a}$. Therefore, $\boldsymbol{a} = I\boldsymbol{a} = (\hat{\mathbf{n}}\hat{\mathbf{n}}^T - \hat{N}^2)\,\boldsymbol{a} = \boldsymbol{a}_\parallel + \boldsymbol{a}_\perp$.

Differential Identities

$$\nabla\times(\nabla\psi) = 0 \tag{C.12}$$

$$\nabla\cdot(\nabla\times\boldsymbol{A}) = 0 \tag{C.13}$$

$$\nabla\cdot(\psi\boldsymbol{A}) = \boldsymbol{A}\cdot\nabla\psi + \psi\nabla\cdot\boldsymbol{A} \tag{C.14}$$

$$\nabla\times(\psi\boldsymbol{A}) = \psi\nabla\times\boldsymbol{A} + \nabla\psi\times\boldsymbol{A} \tag{C.15}$$

$$\nabla(\boldsymbol{A}\cdot\boldsymbol{B}) = (\boldsymbol{A}\cdot\nabla)\boldsymbol{B} + (\boldsymbol{B}\cdot\nabla)\boldsymbol{A} + \boldsymbol{A}\times(\nabla\times\boldsymbol{B}) + \boldsymbol{B}\times(\nabla\times\boldsymbol{A}) \tag{C.16}$$

$$\nabla\cdot(\boldsymbol{A}\times\boldsymbol{B}) = \boldsymbol{B}\cdot(\nabla\times\boldsymbol{A}) - \boldsymbol{A}\cdot(\nabla\times\boldsymbol{B}) \tag{C.17}$$

$$\nabla\times(\boldsymbol{A}\times\boldsymbol{B}) = \boldsymbol{A}(\nabla\cdot\boldsymbol{B}) - \boldsymbol{B}(\nabla\cdot\boldsymbol{A}) + (\boldsymbol{B}\cdot\nabla)\boldsymbol{A} - (\boldsymbol{A}\cdot\nabla)\boldsymbol{B} \tag{C.18}$$

$$\nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} \tag{C.19}$$

$$A_x\nabla B_x + A_y\nabla B_y + A_z\nabla B_z = (\boldsymbol{A}\cdot\nabla)\boldsymbol{B} + \boldsymbol{A}\times(\nabla\times\boldsymbol{B}) \tag{C.20}$$

$$B_x\nabla A_x + B_y\nabla A_y + B_z\nabla A_z = (\boldsymbol{B}\cdot\nabla)\boldsymbol{A} + \boldsymbol{B}\times(\nabla\times\boldsymbol{A}) \tag{C.21}$$

$$(\hat{\mathbf{n}}\times\nabla)\times\boldsymbol{A} = \hat{\mathbf{n}}\times(\nabla\times\boldsymbol{A}) + (\hat{\mathbf{n}}\cdot\nabla)\boldsymbol{A} - \hat{\mathbf{n}}(\nabla\cdot\boldsymbol{A}) \tag{C.22}$$

$$\begin{aligned}\psi(\hat{\mathbf{n}}\cdot\nabla)\boldsymbol{E} - \boldsymbol{E}(\hat{\mathbf{n}}\cdot\nabla\psi) = {}& \big[(\hat{\mathbf{n}}\cdot\nabla)(\psi\boldsymbol{E}) + \hat{\mathbf{n}}\times\big(\nabla\times(\psi\boldsymbol{E})\big) - \hat{\mathbf{n}}\nabla\cdot(\psi\boldsymbol{E})\big] \\ & + \big[\hat{\mathbf{n}}\,\psi\,\nabla\cdot\boldsymbol{E} - (\hat{\mathbf{n}}\times\boldsymbol{E})\times\nabla\psi - \psi\,\hat{\mathbf{n}}\times(\nabla\times\boldsymbol{E}) - (\hat{\mathbf{n}}\cdot\boldsymbol{E})\,\nabla\psi\big]\end{aligned} \tag{C.23}$$

With $\mathbf{r} = x\,\hat{\mathbf{x}} + y\,\hat{\mathbf{y}} + z\,\hat{\mathbf{z}}$, $r = |\mathbf{r}| = \sqrt{x^2+y^2+z^2}$, and the unit vector $\hat{\mathbf{r}} = \mathbf{r}/r$, we have:

$$\nabla r = \hat{\mathbf{r}}\,, \quad \nabla r^2 = 2\mathbf{r}\,, \quad \nabla\frac{1}{r} = -\frac{\hat{\mathbf{r}}}{r^2}\,, \quad \nabla\cdot\mathbf{r} = 3\,, \quad \nabla\times\mathbf{r} = 0\,, \quad \nabla\cdot\hat{\mathbf{r}} = \frac{2}{r} \tag{C.24}$$

Integral Theorems for Closed Surfaces

The theorems involve a volume V surrounded by a closed surface S. The divergence or Gauss' theorem is:

$$\boxed{\int_V \nabla\cdot\boldsymbol{A}\,dV = \oint_S \boldsymbol{A}\cdot\hat{\mathbf{n}}\,dS} \qquad \text{(Gauss' divergence theorem)} \tag{C.25}$$

where $\hat{\mathbf{n}}$ is the *outward* normal to the surface. Green's first and second identities are:

$$\int_V\big[\varphi\nabla^2\psi + \nabla\varphi\cdot\nabla\psi\big]\,dV = \oint_S \varphi\frac{\partial\psi}{\partial n}\,dS \tag{C.26}$$

$$\int_V\big[\varphi\nabla^2\psi - \psi\nabla^2\varphi\big]\,dV = \oint_S\Big(\varphi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\varphi}{\partial n}\Big)\,dS \tag{C.27}$$

where $\dfrac{\partial}{\partial n} = \hat{\mathbf{n}}\cdot\nabla$ is the directional derivative along $\hat{\mathbf{n}}$. Some related theorems are:

$$\int_V \nabla^2\psi\, dV = \oint_S \hat{\mathbf{n}}\cdot\nabla\psi\, dS = \oint_S \frac{\partial\psi}{\partial n}\, dS \tag{C.28}$$

$$\int_V \nabla\psi\, dV = \oint_S \psi\,\hat{\mathbf{n}}\, dS \tag{C.29}$$

$$\int_V \nabla^2\mathbf{A}\, dV = \oint_S (\hat{\mathbf{n}}\cdot\nabla)\mathbf{A}\, dS = \oint_S \frac{\partial\mathbf{A}}{\partial n}\, dS \tag{C.30}$$

$$\oint_S (\hat{\mathbf{n}}\times\nabla)\times\mathbf{A}\, dS = \oint_S \big[\hat{\mathbf{n}}\times(\nabla\times\mathbf{A}) + (\hat{\mathbf{n}}\cdot\nabla)\mathbf{A} - \hat{\mathbf{n}}(\nabla\cdot\mathbf{A})\big]\, dS = 0 \tag{C.31}$$

$$\int_V \nabla\times\mathbf{A}\, dV = \oint_S \hat{\mathbf{n}}\times\mathbf{A}\, dS \tag{C.32}$$

Using Eqs. (C.23) and (C.31), we find:

$$\begin{aligned}&\oint_S \Big(\psi\,\frac{\partial\mathbf{E}}{\partial n} - \mathbf{E}\,\frac{\partial\psi}{\partial n}\Big)\, dS = \\ &\quad = \oint_S \big[\hat{\mathbf{n}}\,\psi\,\nabla\cdot\mathbf{E} - (\hat{\mathbf{n}}\times\mathbf{E})\times\nabla\psi - \psi\,\hat{\mathbf{n}}\times(\nabla\times\mathbf{E}) - (\hat{\mathbf{n}}\cdot\mathbf{E})\,\nabla\psi\big]\, dS\end{aligned} \tag{C.33}$$

The vectorial forms of Green's identities are [1116,1113]:

$$\int_V (\nabla\times\mathbf{A}\cdot\nabla\times\mathbf{B} - \mathbf{A}\cdot\nabla\times\nabla\times\mathbf{B})\, dV = \oint_S \hat{\mathbf{n}}\cdot(\mathbf{A}\times\nabla\times\mathbf{B})\, dS \tag{C.34}$$

$$\int_V (\mathbf{B}\cdot\nabla\times\nabla\times\mathbf{A} - \mathbf{A}\cdot\nabla\times\nabla\times\mathbf{B})\, dV = \oint_S \hat{\mathbf{n}}\cdot(\mathbf{A}\times\nabla\times\mathbf{B} - \mathbf{B}\times\nabla\times\mathbf{A})\, dS \tag{C.35}$$

Integral Theorems for Open Surfaces

Stokes' theorem involves an open surface S and its boundary contour C:

$$\boxed{\int_S \hat{\mathbf{n}}\cdot\nabla\times\mathbf{A}\, dS = \oint_C \mathbf{A}\cdot d\mathbf{l}} \qquad \text{(Stokes' theorem)} \tag{C.36}$$

where $d\mathbf{l}$ is the tangential path length around C. Some related theorems are:

$$\int_S \big[\psi\,\hat{\mathbf{n}}\cdot\nabla\times\mathbf{A} - (\hat{\mathbf{n}}\times\mathbf{A})\cdot\nabla\psi\big]\, dS = \oint_C \psi\mathbf{A}\cdot d\mathbf{l} \tag{C.37}$$

$$\int_S \big[(\nabla\psi)\,\hat{\mathbf{n}}\cdot\nabla\times\mathbf{A} - \big((\hat{\mathbf{n}}\times\mathbf{A})\cdot\nabla\big)\nabla\psi\big]\, dS = \oint_C (\nabla\psi)\mathbf{A}\cdot d\mathbf{l} \tag{C.38}$$

$$\int_S \hat{\mathbf{n}}\times\nabla\psi\, dS = \oint_C \psi\, d\mathbf{l} \tag{C.39}$$

$$\int_S (\hat{\mathbf{n}}\times\nabla)\times\mathbf{A}\, dS = \int_S \big[\hat{\mathbf{n}}\times(\nabla\times\mathbf{A}) + (\hat{\mathbf{n}}\cdot\nabla)\mathbf{A} - \hat{\mathbf{n}}(\nabla\cdot\mathbf{A})\big]\, dS = \oint_C d\mathbf{l}\times\mathbf{A} \tag{C.40}$$

$$\int_S \hat{\mathbf{n}}\, dS = \frac{1}{2}\oint_C \mathbf{r}\times d\mathbf{l} \tag{C.41}$$

Eq. (C.41) is a special case of (C.40). Using Eqs. (C.23) and (C.40) we find:

$$\begin{aligned}&\int_S \Big(\psi\,\frac{\partial\mathbf{E}}{\partial n} - \mathbf{E}\,\frac{\partial\psi}{\partial n}\Big)\, dS + \oint_C \psi\mathbf{E}\times d\mathbf{l} = \\ &\quad = \int_S \big[\hat{\mathbf{n}}\,\psi\,\nabla\cdot\mathbf{E} - (\hat{\mathbf{n}}\times\mathbf{E})\times\nabla\psi - \psi\,\hat{\mathbf{n}}\times(\nabla\times\mathbf{E}) - (\hat{\mathbf{n}}\cdot\mathbf{E})\,\nabla\psi\big]\, dS\end{aligned} \tag{C.42}$$

D. Green's Functions

The Green's functions for the Laplace, Helmholtz, and one-dimensional Helmholtz equations are listed below:

$$\nabla^2 g(\mathbf{r}) = -\delta^{(3)}(\mathbf{r}) \quad\Rightarrow\quad g(\mathbf{r}) = \frac{1}{4\pi r} \tag{D.1}$$

$$(\nabla^2 + k^2)G(\mathbf{r}) = -\delta^{(3)}(\mathbf{r}) \quad\Rightarrow\quad G(\mathbf{r}) = \frac{e^{-jkr}}{4\pi r} \tag{D.2}$$

$$(\partial_z^2 + \beta^2)g(z) = -\delta(z) \quad\Rightarrow\quad g(z) = \frac{e^{-j\beta|z|}}{2j\beta} \tag{D.3}$$

where $r = |\mathbf{r}|$. Eqs. (D.2) and (D.3) are appropriate for describing *outgoing* waves. We considered other versions of (D.3) in Sec. 21.3. A more general identity satisfied by the Green's function $g(\mathbf{r})$ of Eq. (D.1) is as follows (for a proof, see Refs. [111,112]):

$$\partial_i\partial_j g(\mathbf{r}) = -\frac{1}{3}\delta_{ij}\,\delta^{(3)}(\mathbf{r}) + \frac{3x_ix_j - r^2\delta_{ij}}{r^4}\,g(\mathbf{r}) \qquad i,j = 1,2,3 \tag{D.4}$$

where $\partial_i = \partial/\partial x_i$ and x_i stands for any of x, y, z. By summing the i, j indices, Eq. (D.4) reduces to (D.1). Using this identity, we find for the Green's function $G(\mathbf{r}) = e^{-jkr}/4\pi r$:

$$\partial_i\partial_j G(\mathbf{r}) = -\frac{1}{3}\delta_{ij}\,\delta^{(3)}(\mathbf{r}) + \left[\Big(jk + \frac{1}{r}\Big)\frac{3x_ix_j - r^2\delta_{ij}}{r^3} - k^2\,\frac{x_ix_j}{r^2}\right]G(\mathbf{r}) \tag{D.5}$$

This reduces to Eq. (D.2) upon summing the indices. For any fixed vector $\mathbf{p}$, Eq. (D.5) is equivalent to the vectorial identity:

$$\nabla\times\nabla\times\big[\mathbf{p}\,G(\mathbf{r})\big] = \frac{2}{3}\mathbf{p}\,\delta^{(3)}(\mathbf{r}) + \left[\Big(jk + \frac{1}{r}\Big)\frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\mathbf{p}) - \mathbf{p}}{r^2} + k^2\,\hat{\mathbf{r}}\times(\mathbf{p}\times\hat{\mathbf{r}})\right]G(\mathbf{r}) \tag{D.6}$$

The second term on the right is simply the left-hand side evaluated at points away from the origin, thus, we may write:

$$\nabla\times\nabla\times\big[\mathbf{p}\,G(\mathbf{r})\big] = \frac{2}{3}\mathbf{p}\,\delta^{(3)}(\mathbf{r}) + \Big[\nabla\times\nabla\times\big[\mathbf{p}\,G(\mathbf{r})\big]\Big]_{\mathbf{r}\neq 0} \tag{D.7}$$

Then, Eq. (D.7) implies the following integrated identity, where $\boldsymbol{\nabla}$ is with respect to $\mathbf{r}$:

$$\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\int_V \boldsymbol{P}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\,dV' = \frac{2}{3}\,\boldsymbol{P}(\mathbf{r})+\int_V\Big[\boldsymbol{\nabla}\times\boldsymbol{\nabla}\times\big[\boldsymbol{P}(\mathbf{r}')\,G(\mathbf{r}-\mathbf{r}')\big]\Big]_{\mathbf{r}'\neq\mathbf{r}}dV' \tag{D.8}$$

and $\mathbf{r}$ is assumed to lie within V. If $\mathbf{r}$ is outside V, then the term $2\boldsymbol{P}(\mathbf{r})/3$ is absent.

Technically, the integrals in (D.8) are *principal-value* integrals, that is, the limits as $\delta\to 0$ of the integrals over $V-V_\delta(\mathbf{r})$, where $V_\delta(\mathbf{r})$ is an excluded small sphere of radius δ centered about $\mathbf{r}$. The $2\boldsymbol{P}(\mathbf{r})/3$ term has a different form if the excluded volume $V_\delta(\mathbf{r})$ has shape other than a sphere or a cube. See Refs. [1153,460,472,598] and [106–110] for the definitions and properties of such principal value integrals.

Another useful result is the so-called *Weyl representation* or plane-wave-spectrum representation [22,26,1153,27,515] of the outgoing Helmholtz Green's function $G(\mathbf{r})$:

$$\boxed{G(\mathbf{r})=\frac{e^{-jkr}}{4\pi r} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{-j(k_x x+k_y y)}\,e^{-jk_z|z|}}{2jk_z}\,\frac{dk_x\,dk_y}{(2\pi)^2}} \tag{D.9}$$

where $k_z^2 = k^2 - k_\perp^2$, with $k_\perp = \sqrt{k_x^2+k_y^2}$. In order to correspond to either outgoing waves or decaying evanescent waves, k_z must be defined more precisely as follows:

$$k_z = \begin{cases} \sqrt{k^2-k_\perp^2}\,, & \text{if}\quad k_\perp \le k\,, \quad \text{(propagating modes)}\\ -j\sqrt{k_\perp^2-k^2}\,, & \text{if}\quad k_\perp > k\,, \quad \text{(evanescent modes)}\end{cases} \tag{D.10}$$

The propagating modes are important in radiation problems and conventional imaging systems, such as Fourier optics [1156]. The evanescent modes are important in the new subject of *near-field optics*, in which objects can be probed and imaged at nanometer scales improving the resolution of optical microscopy by factors of ten. Some near-field optics references are [494–514].

To prove (D.9), we consider the two-dimensional spatial Fourier transform of $G(\mathbf{r})$ and its inverse. Indicating explicitly the dependence on the coordinates x,y,z, we have:

$$\begin{aligned} g(k_x,k_y,z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x,y,z)\,e^{j(k_x x+k_y y)}\,dx\,dy = \frac{e^{-jk_z|z|}}{2jk_z}\\ G(x,y,z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(k_x,k_y,z)\,e^{-j(k_x x+k_y y)}\,\frac{dk_x\,dk_y}{(2\pi)^2}\end{aligned} \tag{D.11}$$

Writing $\delta^{(3)}(\mathbf{r})=\delta(x)\,\delta(y)\,\delta(z)$ and using the inverse Fourier transform:

$$\delta(x)\,\delta(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-j(k_x x+k_y y)}\,\frac{dk_x\,dk_y}{(2\pi)^2}\,,$$

we find from Eq. (D.2) that $g(k_x,k_y,z)$ must satisfy the one-dimensional Helmholtz Green's function equation (D.3), with $k_z^2 = k^2 - k_x^2 - k_y^2 = k^2 - k_\perp^2$, that is,

$$(\partial_z^2 + k_z^2)\,g(k_x,k_y,z) = -\delta(z) \tag{D.12}$$

whose outgoing/evanescent solution is $g(k_x,k_y,z) = e^{-jk_z|z|}/2jk_z$.

A more direct proof of (D.9) is to use cylindrical coordinates, $k_x = k_\perp\cos\psi$, $k_y = k_\perp\sin\psi$, $x=\rho\cos\phi$, $y=\rho\sin\phi$, where $k_\perp^2 = k_x^2+k_y^2$ and $\rho^2 = x^2+y^2$. It follows that $k_x x + k_y y = k_\perp\rho\cos(\phi-\psi)$. Setting $dx\,dy = \rho\,d\rho\,d\phi = r\,dr\,d\phi$, the latter following from $r^2 = \rho^2+z^2$, we obtain from Eq. (D.11) after replacing $\rho = \sqrt{r^2-z^2}$:

$$\begin{aligned} g(k_x,k_y,z) &= \iint \frac{e^{-jkr}}{4\pi r}\,e^{j(k_x x+k_y y)}\,dx\,dy = \iint \frac{e^{-jkr}}{4\pi r}\,e^{jk_\perp\rho\cos(\phi-\psi)}\,r\,dr\,d\phi\\ &= \frac{1}{2}\int_{|z|}^{\infty} dr\,e^{-jkr}\int_0^{2\pi}\frac{d\phi}{2\pi}\,e^{jk_\perp\rho\cos(\phi-\psi)} = \frac{1}{2}\int_{|z|}^{\infty} dr\,e^{-jkr}\,J_0\big(k_\perp\sqrt{r^2-z^2}\big)\end{aligned}$$

where we used the integral representation (17.9.2) of the Bessel function $J_0(x)$. Looking up the last integral in the table of integrals [1268], we find:

$$\boxed{g(k_x,k_y,z) = \frac{1}{2}\int_{|z|}^{\infty} dr\,e^{-jkr}\,J_0\big(k_\perp\sqrt{r^2-z^2}\big) = \frac{e^{-jk_z|z|}}{2jk_z}} \tag{D.13}$$

where k_z must be defined exactly as in Eq. (D.10). A direct consequence of Eq. (D.11) and the even-ness of $G(\mathbf{r})$ in $\mathbf{r}$ and of $g(k_x,k_y,z)$ in k_x,k_y, is the following result:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-j(k_x x'+k_y y')}\,G(\mathbf{r}-\mathbf{r}')\,dx'dy' = e^{-j(k_x x+k_y y)}\,\frac{e^{-jk_z|z-z'|}}{2jk_z} \tag{D.14}$$

One can also show the integral:

$$\int_0^{\infty} e^{-jk'_z z'}\,\frac{e^{-jk_z|z-z'|}}{2jk_z}\,dz' = \begin{cases}\dfrac{e^{-jk'_z z}}{k'^2_z-k_z^2} - \dfrac{e^{-jk_z z}}{2k_z(k'_z-k_z)}\,, & \text{for}\quad z\ge 0\\[2ex] -\dfrac{e^{jk_z z}}{2k_z(k'_z+k_z)}\,, & \text{for}\quad z<0\end{cases} \tag{D.15}$$

The proof is obtained by splitting the integral over the sub-intervals $[0,z]$ and $[z,\infty)$. To handle the limits at infinity, k'_z must be assumed to be slightly lossy, that is, $k'_z = \beta_z - j\alpha_z$, with $\alpha_z > 0$. Eqs. (D.14) and (D.15) can be combined into:

$$\int_{V_+} e^{-j\boldsymbol{k}'\cdot\mathbf{r}'}\,G(\mathbf{r}-\mathbf{r}')\,dV' = \begin{cases}\dfrac{e^{-j\boldsymbol{k}'\cdot\mathbf{r}}}{k'^2-k^2} - \dfrac{e^{-j\boldsymbol{k}\cdot\mathbf{r}}}{2k_z(k'_z-k_z)}\,, & \text{for}\quad z\ge 0\\[2ex] -\dfrac{e^{-j\boldsymbol{k}_-\cdot\mathbf{r}}}{2k_z(k'_z+k_z)}\,, & \text{for}\quad z<0\end{cases} \tag{D.16}$$

where V_+ is the half-space $z\ge 0$, and $\boldsymbol{k}$, $\boldsymbol{k}_-$, $\boldsymbol{k}'$ are wave-vectors with the same k_x,k_y components, but different k_zs:

$$\begin{aligned}\boldsymbol{k} &= k_x\,\hat{\mathbf{x}} + k_y\,\hat{\mathbf{y}} + k_z\,\hat{\mathbf{z}}\\ \boldsymbol{k}_- &= k_x\,\hat{\mathbf{x}} + k_y\,\hat{\mathbf{y}} - k_z\,\hat{\mathbf{z}}\\ \boldsymbol{k}' &= k_x\,\hat{\mathbf{x}} + k_y\,\hat{\mathbf{y}} + k'_z\,\hat{\mathbf{z}}\end{aligned} \tag{D.17}$$

where we note that $k'^2 - k^2 = (k_x^2+k_y^2+k'^2_z) - (k_x^2+k_y^2+k_z^2) = k'^2_z - k_z^2$.

The Green's function results (D.8)–(D.17) are used in the discussion of the Ewald-Oseen extinction theorem in Sec. 14.6.

A related Weyl-type representation is obtained by differentiating Eq. (D.9) with respect to z. Assuming that $z \geq 0$ and interchanging differentiation and integration (and multiplying by -2), we obtain the identity:

$$\boxed{-2\frac{\partial}{\partial z}\left(\frac{e^{-jkr}}{4\pi r}\right) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} e^{-jk_x x}\, e^{-jk_y y} e^{-jk_z z}\, \frac{dk_x\, dk_y}{(2\pi)^2}}\,, \quad z \geq 0 \tag{D.18}$$

This just means that the left-hand side is the two-dimensional inverse Fourier transform of $e^{-jk_z z}$ with k_z given by Eq. (D.10). Replacing $\mathbf{r}$ by $\mathbf{r}-\mathbf{r}'$, and r by $R = |\mathbf{r}-\mathbf{r}'|$, and noting that $\partial_{z'} = -\partial_z$, we also obtain:

$$\boxed{2\frac{\partial}{\partial z'}\left(\frac{e^{-jkR}}{4\pi R}\right) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} e^{-jk_x(x-x')}\, e^{-jk_y(y-y')} e^{-jk_z(z-z')}\, \frac{dk_x\, dk_y}{(2\pi)^2}}\,, \quad z \geq z' \tag{D.19}$$

This result establishes the equivalence between the Kirchhoff-Fresnel diffraction formula and the plane-wave spectrum representation as discussed in Sec. 17.17. For the vector diffraction case, we also need the derivatives of G with respect to the transverse coordinates x, y. Differentiating (D.9) with respect to x (or with respect to y), we have:

$$\boxed{-2\frac{\partial}{\partial x}\left(\frac{e^{-jkr}}{4\pi r}\right) = \int_{-\infty}^{\infty}\!\int_{-\infty}^{\infty} \frac{k_x}{k_z}\, e^{-jk_x x}\, e^{-jk_y y} e^{-jk_z z}\, \frac{dk_x\, dk_y}{(2\pi)^2}}\,, \quad z \geq 0 \tag{D.20}$$

E. Coordinate Systems

The definitions of cylindrical and spherical coordinates were given in Sec. 14.8. The expressions of the gradient, divergence, curl, Laplacian operators, and delta functions are given below in cartesian, cylindrical, and spherical coordinates.

Cartesian Coordinates

$$\nabla\psi = \hat{\mathbf{x}}\,\frac{\partial\psi}{\partial x} + \hat{\mathbf{y}}\,\frac{\partial\psi}{\partial y} + \hat{\mathbf{z}}\,\frac{\partial\psi}{\partial z}$$

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}$$

$$\nabla\cdot\boldsymbol{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}$$

$$\nabla\times\boldsymbol{A} = \hat{\mathbf{x}}\left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right) + \hat{\mathbf{y}}\left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right) + \hat{\mathbf{z}}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{E.1}$$

$$= \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial_x & \partial_y & \partial_z \\ A_x & A_y & A_z \end{vmatrix}$$

$$\delta^{(3)}(\mathbf{r}-\mathbf{r}') = \delta(x-x')\,\delta(y-y')\,\delta(z-z')$$

Cylindrical Coordinates

$$\nabla\psi = \hat{\boldsymbol{\rho}}\,\frac{\partial\psi}{\partial\rho} + \hat{\boldsymbol{\phi}}\,\frac{1}{\rho}\,\frac{\partial\psi}{\partial\phi} + \hat{\mathbf{z}}\,\frac{\partial\psi}{\partial z} \tag{E.2a}$$

$$\nabla^2\psi = \frac{1}{\rho}\,\frac{\partial}{\partial\rho}\left(\rho\,\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\,\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2} \tag{E.2b}$$

$$\nabla\cdot\boldsymbol{A} = \frac{1}{\rho}\,\frac{\partial(\rho A_\rho)}{\partial\rho} + \frac{1}{\rho}\,\frac{\partial A_\phi}{\partial\phi} + \frac{\partial A_z}{\partial z} \tag{E.2c}$$

$$\nabla\times\boldsymbol{A} = \hat{\boldsymbol{\rho}}\left(\frac{1}{\rho}\,\frac{\partial A_z}{\partial\phi} - \frac{\partial A_\phi}{\partial z}\right) + \hat{\boldsymbol{\phi}}\left(\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial\rho}\right) + \hat{\mathbf{z}}\,\frac{1}{\rho}\left(\frac{\partial(\rho A_\phi)}{\partial\rho} - \frac{\partial A_\rho}{\partial\phi}\right) \tag{E.2d}$$

$$\delta^{(3)}(\mathbf{r}-\mathbf{r}') = \frac{1}{\rho}\,\delta(\rho-\rho')\,\delta(\phi-\phi')\,\delta(z-z') \tag{E.2e}$$

Spherical Coordinates

$$\nabla\psi = \hat{\mathbf{r}}\,\frac{\partial\psi}{\partial r} + \hat{\boldsymbol{\theta}}\,\frac{1}{r}\,\frac{\partial\psi}{\partial\theta} + \hat{\boldsymbol{\phi}}\,\frac{1}{r\sin\theta}\,\frac{\partial\psi}{\partial\phi} \tag{E.3a}$$

$$\nabla^2\psi = \frac{1}{r^2}\,\frac{\partial}{\partial r}\left(r^2\,\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\,\frac{\partial^2\psi}{\partial\phi^2} \tag{E.3b}$$

$$\nabla\cdot\boldsymbol{A} = \frac{1}{r^2}\,\frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r\sin\theta}\,\frac{\partial(\sin\theta A_\theta)}{\partial\theta} + \frac{1}{r\sin\theta}\,\frac{\partial A_\phi}{\partial\phi} \tag{E.3c}$$

$$\nabla\times\boldsymbol{A} = \hat{\mathbf{r}}\,\frac{1}{r\sin\theta}\left(\frac{\partial(\sin\theta A_\phi)}{\partial\theta} - \frac{\partial A_\theta}{\partial\phi}\right) + \hat{\boldsymbol{\theta}}\,\frac{1}{r}\left(\frac{1}{\sin\theta}\,\frac{\partial A_r}{\partial\phi} - \frac{\partial(rA_\phi)}{\partial r}\right) \tag{E.3d}$$

$$+ \hat{\boldsymbol{\phi}}\,\frac{1}{r}\left(\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial\theta}\right)$$

$$\delta^{(3)}(\mathbf{r}-\mathbf{r}') = \frac{1}{r^2\sin\theta}\,\delta(r-r')\,\delta(\theta-\theta')\,\delta(\phi-\phi') \tag{E.3e}$$

Transformations Between Coordinate Systems

A vector $\boldsymbol{A}$ can be expressed component-wise in the three coordinate systems as:

$$\begin{aligned} \boldsymbol{A} &= \hat{\mathbf{x}}\,A_x + \hat{\mathbf{y}}\,A_y + \hat{\mathbf{z}}\,A_z \\ &= \hat{\boldsymbol{\rho}}\,A_\rho + \hat{\boldsymbol{\phi}}\,A_\phi + \hat{\mathbf{z}}\,A_z \\ &= \hat{\mathbf{r}}\,A_r + \hat{\boldsymbol{\theta}}\,A_\theta + \hat{\boldsymbol{\phi}}\,A_\phi \end{aligned} \tag{E.4}$$

The components in one coordinate system can be expressed in terms of the components of another by using the following relationships between the unit vectors, which were also given in Eqs. (14.8.1)–(14.8.3):

$$\begin{aligned} x &= \rho\cos\phi & \hat{\boldsymbol{\rho}} &= \hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi & \hat{\mathbf{x}} &= \hat{\boldsymbol{\rho}}\cos\phi - \hat{\boldsymbol{\phi}}\sin\phi \\ y &= \rho\sin\phi & \hat{\boldsymbol{\phi}} &= -\hat{\mathbf{x}}\sin\phi + \hat{\mathbf{y}}\cos\phi & \hat{\mathbf{y}} &= \hat{\boldsymbol{\rho}}\sin\phi + \hat{\boldsymbol{\phi}}\cos\phi \end{aligned} \tag{E.5}$$

$$\begin{aligned} \rho &= r\sin\theta & \hat{\mathbf{r}} &= \hat{\mathbf{z}}\cos\theta + \hat{\boldsymbol{\rho}}\sin\theta & \hat{\mathbf{z}} &= \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta \\ z &= r\cos\theta & \hat{\boldsymbol{\theta}} &= -\hat{\mathbf{z}}\sin\theta + \hat{\boldsymbol{\rho}}\cos\theta & \hat{\boldsymbol{\rho}} &= \hat{\mathbf{r}}\sin\theta + \hat{\boldsymbol{\theta}}\cos\theta \end{aligned} \tag{E.6}$$

$$\begin{aligned} x &= r\sin\theta\cos\phi & \hat{\mathbf{r}} &= \hat{\mathbf{x}}\cos\phi\sin\theta + \hat{\mathbf{y}}\sin\phi\sin\theta + \hat{\mathbf{z}}\cos\theta \\ y &= r\sin\theta\sin\phi & \hat{\boldsymbol{\theta}} &= \hat{\mathbf{x}}\cos\phi\cos\theta + \hat{\mathbf{y}}\sin\phi\cos\theta - \hat{\mathbf{z}}\sin\theta \\ z &= r\cos\theta & \hat{\boldsymbol{\phi}} &= -\hat{\mathbf{x}}\sin\phi + \hat{\mathbf{y}}\cos\phi \end{aligned} \tag{E.7}$$

$$\begin{aligned} \hat{\mathbf{x}} &= \hat{\mathbf{r}}\sin\theta\cos\phi + \hat{\boldsymbol{\theta}}\cos\theta\cos\phi - \hat{\boldsymbol{\phi}}\sin\phi \\ \hat{\mathbf{y}} &= \hat{\mathbf{r}}\sin\theta\sin\phi + \hat{\boldsymbol{\theta}}\cos\theta\sin\phi + \hat{\boldsymbol{\phi}}\cos\phi \\ \hat{\mathbf{z}} &= \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta \end{aligned} \tag{E.8}$$

For example, to express the spherical components A_θ, A_ϕ in terms of the cartesian components, we proceed as follows:

$$A_\theta = \hat{\boldsymbol{\theta}}\cdot\mathbf{A} = \hat{\boldsymbol{\theta}}\cdot(\hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z) = (\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{x}})\,A_x + (\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{y}})\,A_y + (\hat{\boldsymbol{\theta}}\cdot\hat{\mathbf{z}})\,A_z$$

$$A_\phi = \hat{\boldsymbol{\phi}}\cdot\mathbf{A} = \hat{\boldsymbol{\phi}}\cdot(\hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z) = (\hat{\boldsymbol{\phi}}\cdot\hat{\mathbf{x}})\,A_x + (\hat{\boldsymbol{\phi}}\cdot\hat{\mathbf{y}})\,A_y + (\hat{\boldsymbol{\phi}}\cdot\hat{\mathbf{z}})\,A_z$$

The dot products can be read off Eq. (E.7), resulting in:

$$\begin{aligned} A_\theta &= \cos\phi\cos\theta A_x + \sin\phi\cos\theta A_y - \sin\theta A_z \\ A_\phi &= -\sin\phi A_x + \cos\phi A_y \end{aligned} \tag{E.9}$$

Similarly, using Eq. (E.6) the cylindrical components A_ρ, A_z can be expressed in terms of spherical components as:

$$\begin{aligned} A_\rho &= \hat{\boldsymbol{\rho}}\cdot\mathbf{A} = \hat{\boldsymbol{\rho}}\cdot(\hat{\mathbf{r}}A_r + \hat{\boldsymbol{\theta}}A_\theta + \hat{\boldsymbol{\phi}}A_\phi) = \sin\theta A_r + \cos\theta A_\theta \\ A_z &= \hat{\mathbf{z}}\cdot\mathbf{A} = \hat{\mathbf{z}}\cdot(\hat{\mathbf{r}}A_r + \hat{\boldsymbol{\theta}}A_\theta + \hat{\boldsymbol{\phi}}A_\phi) = \cos\theta A_r - \cos\theta A_\theta \end{aligned} \tag{E.10}$$

F. Fresnel, Exponential, Sine, and Cosine Integrals

The Fresnel functions $C(x)$ and $S(x)$ are defined by [1267]:

$$C(x) = \int_0^x \cos\left(\frac{\pi}{2}t^2\right)dt\,, \quad S(x) = \int_0^x \sin\left(\frac{\pi}{2}t^2\right)dt \tag{F.1}$$

They may be combined into the complex function:

$$\boxed{\mathcal{F}(x) = C(x) - jS(x) = \int_0^x e^{-j(\pi/2)t^2}\,dt} \tag{F.2}$$

$C(x)$, $S(x)$, and $\mathcal{F}(x)$ are *odd* functions of x and have the asymptotic values:

$$C(\infty) = S(\infty) = \frac{1}{2}\,, \quad \mathcal{F}(\infty) = \frac{1-j}{2} \tag{F.3}$$

At $x = 0$, we have $\mathcal{F}(0) = 0$ and $\mathcal{F}'(0) = 1$, so that the Taylor series approximation is $\mathcal{F}(x) \simeq x$, for small x. The asymptotic expansions of $C(x)$, $S(x)$, and $\mathcal{F}(x)$ are for large positive x:

$$\begin{aligned} \mathcal{F}(x) &= \frac{1-j}{2} + \frac{j}{\pi x}\,e^{-j\pi x^2/2} \\ C(x) &= \frac{1}{2} + \frac{1}{\pi x}\,\sin\left(\frac{\pi}{2}x^2\right) \\ S(x) &= \frac{1}{2} - \frac{1}{\pi x}\,\cos\left(\frac{\pi}{2}x^2\right) \end{aligned} \tag{F.4}$$

Associated with $C(x)$ and $S(x)$ are the type-2 Fresnel integrals:

$$C_2(x) = \int_0^x \frac{\cos t}{\sqrt{2\pi t}}\,dt\,, \quad S_2(x) = \int_0^x \frac{\sin t}{\sqrt{2\pi t}}\,dt \tag{F.5}$$

They are combined into the complex function:

$$\mathcal{F}_2(x) = C_2(x) - jS_2(x) = \int_0^x \frac{e^{-jt}}{\sqrt{2\pi t}}\,dt \tag{F.6}$$

The two types are related by, if $x \ge 0$:

$$C(x) = C_2\left(\frac{\pi}{2}x^2\right), \quad S(x) = S_2\left(\frac{\pi}{2}x^2\right), \quad \mathcal{F}(x) = \mathcal{F}_2\left(\frac{\pi}{2}x^2\right) \tag{F.7}$$

and if $x < 0$, we set $\mathcal{F}(x) = -\mathcal{F}(-x) = -\mathcal{F}_2(\pi x^2/2)$.

The Fresnel function $\mathcal{F}_2(x)$ can be evaluated numerically using Boersma's approximation [1132], which achieves a maximum error of 10^{-9} over all x. The algorithm approximates the function $\mathcal{F}_2(x)$ as follows:

$$\mathcal{F}_2(x) = \begin{cases} e^{-jx}\sqrt{\dfrac{x}{4}}\displaystyle\sum_{n=0}^{11}(a_n + jb_n)\left(\frac{x}{4}\right)^n, & \text{if} \quad 0 \le x \le 4 \\ \dfrac{1-j}{2} + e^{-jx}\sqrt{\dfrac{4}{x}}\displaystyle\sum_{n=0}^{11}(c_n + jd_n)\left(\frac{4}{x}\right)^n, & \text{if} \quad x > 4 \end{cases} \tag{F.8}$$

where the coefficients a_n, b_n, c_n, d_n are given in [1132]. Consistency with the small- and large-x expansions of $\mathcal{F}(x)$ requires that $a_0 + jb_0 = \sqrt{8/\pi}$ and $c_0 + jd_0 = j/\sqrt{8\pi}$. We have implemented Eq. (F.8) with the MATLAB function `fcs2`:

```
F2 = fcs2(x);          % Fresnel integrals F2(x) = C2(x)-jS2(x)
```

The ordinary Fresnel integral $\mathcal{F}(x)$ can be computed with the help of Eq. (F.7). The MATLAB function `fcs` calculates $\mathcal{F}(x)$ for any vector of values x by calling `fcs2`:

```
F = fcs(x);          % Fresnel integrals F(x) = C(x)-jS(x)
```

In calculating the radiation patterns of pyramidal horns, it is desired to calculate a Fresnel diffraction integral of the type:

$$\boxed{F_0(\nu,\sigma) = \int_{-1}^{1} e^{j\pi\nu\xi}\,e^{-j(\pi/2)\sigma^2\xi^2}\,d\xi} \tag{F.9}$$

Making the variable change $t = \sigma\xi - \nu/\sigma$, this integral can be computed in terms of the Fresnel function $\mathcal{F}(x) = C(x) - jS(x)$ as follows:

$$\boxed{F_0(\nu,\sigma) = \frac{1}{\sigma}\, e^{j(\pi/2)(\nu^2/\sigma^2)}\left[\mathcal{F}\left(\frac{\nu}{\sigma}+\sigma\right) - \mathcal{F}\left(\frac{\nu}{\sigma}-\sigma\right)\right]} \tag{F.10}$$

where we also used the oddness of $\mathcal{F}(x)$. The value of Eq. (F.9) at $\nu = 0$ is:

$$\boxed{F_0(0,\sigma) = \frac{1}{\sigma}\left[\mathcal{F}(\sigma) - \mathcal{F}(-\sigma)\right] = 2\,\frac{\mathcal{F}(\sigma)}{\sigma}} \tag{F.11}$$

Eq. (F.10) assumes that $\sigma \neq 0$. If $\sigma = 0$, the integral (F.9) reduces to the sinc function:

$$\boxed{F_0(\nu,0) = 2\,\frac{\sin(\pi\nu)}{\pi\nu}} \tag{F.12}$$

From either (F.11) or (F.12), we find $F_0(0,0) = 2$. A related integral that is also required in the theory of horns is the following:

$$\boxed{F_1(\nu,\sigma) = \int_{-1}^{1}\cos\left(\frac{\pi\xi}{2}\right) e^{j\pi\nu\xi}\, e^{-j(\pi/2)\sigma^2\xi^2}\, d\xi} \tag{F.13}$$

Writing $\cos(\pi\xi/2) = (e^{j\pi\xi/2} + e^{-j\pi\xi/2})/2$, the integral $F_1(\nu,s)$ can be expressed in terms of $F_0(\nu,\sigma)$ as follows:

$$\boxed{F_1(\nu,\sigma) = \frac{1}{2}\left[F_0(\nu+0.5,\sigma) + F_0(\nu-0.5,\sigma)\right]} \tag{F.14}$$

It can be verified easily that $F_0(0.5,\sigma) = F_0(-0.5,\sigma)$, therefore, the value of $F_1(\nu,\sigma)$ at $\nu = 0$ will be given by:

$$\boxed{F_1(0,\sigma) = F_0(0.5,\sigma) = \frac{1}{\sigma}\, e^{j\pi/(8\sigma^2)}\left[\mathcal{F}\left(\frac{1}{2\sigma}+\sigma\right) - \mathcal{F}\left(\frac{1}{2\sigma}-\sigma\right)\right]} \tag{F.15}$$

Using the asymptotic expansion (F.4), we find the expansion valid for small σ:

$$\mathcal{F}\left(\frac{1}{2\sigma}\pm\sigma\right) = \frac{1-j}{2} \mp \frac{2\sigma}{\pi}\, e^{-j\pi/(8\sigma^2)}\,, \quad \text{for small } \sigma \tag{F.16}$$

For $\sigma = 0$, the integral $F_1(\nu,\sigma)$ reduces to the double-sinc function:

$$\begin{aligned} F_1(\nu,0) &= \int_{-1}^{1}\cos\left(\frac{\pi\xi}{2}\right) e^{j\pi\nu\xi}\, d\xi = \frac{1}{2}\left[F_0(\nu+0.5,0) + F_0(\nu-0.5,0)\right] \\ &= \frac{\sin\big(\pi(\nu+0.5)\big)}{\pi(\nu+0.5)} + \frac{\sin\big(\pi(\nu-0.5)\big)}{\pi(\nu-0.5)} = \frac{4}{\pi}\,\frac{\cos(\pi\nu)}{1-4\nu^2} \end{aligned} \tag{F.17}$$

From either Eq. (F.16) or (F.17), we find $F_1(0,0) = 4/\pi$.

The MATLAB function `diffint` can be used to evaluate both Eq. (F.9) and (F.13) for any vector of values ν and any vector of positive numbers σ, including $\sigma = 0$. It calls `fcs` to evaluate the diffraction integral (F.9) according to Eq. (F.10). Its usage is:

```
F0 = diffint(v,sigma,0);          % diffraction integral F0(ν,σ), Eq. (F.9)
F1 = diffint(v,sigma,1);          % diffraction integral F1(ν,σ), Eq. (F.13)
```

The vectors `v,sigma` can be entered either as rows or columns, but the result will be a matrix of size `length(v) x length(sigma)`. The integral $F_0(\nu,\sigma)$ can also be calculated by the simplified call:

```
F0 = diffint(v,sigma);            % diffraction integral F0(ν,σ), Eq. (F.9)
```

Actually, the most general syntax of `diffint` is as follows:

```
F = diffint(v,sigma,a,c1,c2);     % diffraction integral F(ν,σ,a), Eq. (F.18)
```

It evaluates the more general integral:

$$\boxed{F(\nu,\sigma,a) = \int_{c_1}^{c_2}\cos\left(\frac{\pi\xi a}{2}\right) e^{j\pi\nu\xi}\, e^{-j(\pi/2)\sigma^2\xi^2}\, d\xi} \tag{F.18}$$

For $a = 0$, we have:

$$\boxed{F(\nu,\sigma,0) = \frac{1}{\sigma}\, e^{j(\pi/2)(\nu^2/\sigma^2)}\left[\mathcal{F}\left(\frac{\nu}{\sigma}-\sigma c_1\right) - \mathcal{F}\left(\frac{\nu}{\sigma}-\sigma c_2\right)\right]} \tag{F.19}$$

For $a \neq 0$, we can express $F(\nu,\sigma,a)$ in terms of $F(\nu,\sigma,0)$:

$$\boxed{F(\nu,\sigma,a) = \frac{1}{2}\left[F(\nu+0.5a,\sigma,0) + F(\nu-0.5a,\sigma,0)\right]} \tag{F.20}$$

For $a = 0$ and $\sigma = 0$, $F(\nu,\sigma,a)$ reduces to the complex sinc function:

$$\boxed{F(\nu,0,0) = \frac{e^{j\pi\nu c_2} - e^{j\pi\nu c_1}}{j\pi\nu} = (c_2-c_1)\,\frac{\sin\big(\pi(c_2-c_1)\nu/2\big)}{\pi(c_2-c_1)\nu/2}\, e^{j\pi(c_2+c_1)\nu/2}} \tag{F.21}$$

Stationary Phase Approximation

The Fresnel integrals find also application in the the stationary-phase approximation for evaluating integrals. The approximation can be stated as follows:

$$\boxed{\int_{-\infty}^{\infty} f(x)\, e^{j\phi(x)}\, dx \simeq \sqrt{\frac{2\pi j}{\phi''(x_0)}}\, f(x_0)\, e^{j\phi(x_0)}} \tag{F.22}$$

where x_0 is a stationary point of the phase $\phi(x)$, that is, the solution of $\phi'(x_0) = 0$, where for simplicity we assume that there is only one such point (otherwise, one has a sum of terms like (F.22), one for each solution of $\phi'(x) = 0$). Eq. (F.22) is obtained by expanding $\phi(x)$ in Taylor series about the stationary point $x = x_0$ and keeping only up to the quadratic term:

$$\phi(x) \simeq \phi(x_0) + \phi'(x_0)(x-x_0) + \frac{1}{2}\phi''(x_0)(x-x_0)^2 = \phi(x_0) + \frac{1}{2}\phi''(x_0)(x-x_0)^2$$

Making this approximation in the integral and assuming that $f(x)$ is slowly varying in the neighborhood of x_0, we may replace $f(x)$ by its value at x_0:

$$\int_{-\infty}^{\infty} f(x)\, e^{j\phi(x)}\,dx \simeq \int_{-\infty}^{\infty} f(x_0)\, e^{j\left(\phi(x_0)+\phi''(x_0)(x-x_0)^2/2\right)}\,dx$$

$$= f(x_0)\, e^{j\phi(x_0)} \int_{-\infty}^{\infty} e^{j\phi''(x_0)(x-x_0)^2/2}dx$$

The last integral can be reduced to the complex Fresnel integral by the change of variables $(x-x_0)=\sqrt{\pi/\phi''(x_0)}\,u$:

$$\int_{-\infty}^{\infty} e^{j\phi''(x_0)(x-x_0)^2/2}dx = \sqrt{\frac{\pi}{\phi''(x_0)}}\int_{-\infty}^{\infty} e^{j\pi u^2/2}du = \sqrt{\frac{\pi}{\phi''(x_0)}}\left[\mathcal{F}(\infty)-\mathcal{F}(-\infty)\right]^*$$

Using $\left[\mathcal{F}(\infty)-\mathcal{F}(-\infty)\right]^* = 2\mathcal{F}^*(\infty) = 1+j = \sqrt{2j}$, we obtain

$$\int_{-\infty}^{\infty} e^{j\phi''(x_0)(x-x_0)^2/2}dx = \sqrt{\frac{2\pi j}{\phi''(x_0)}}$$

Normally, the phase depends on a positive parameter λ in the form $\phi(x)=\lambda\theta(x)$, and the stationary-phase approximation is justified in the limit $\lambda\to\infty$.

Exponential, Sine, and Cosine Integrals

Several antenna calculations, such as mutual impedances and directivities, can be reduced to the exponential integral, which is defined as follows [1267]:

$$E_1(z)=\int_z^{\infty}\frac{e^{-u}}{u}\,du = e^{-z}\int_0^{\infty}\frac{e^{-t}}{z+t}\,dt \qquad \text{(exponential integral)} \tag{F.23}$$

where z is a complex number with phase restricted such that $|\arg z|<\pi$. This range allows pure imaginary z's. The built-in MATLAB function `expint` evaluates $E_1(z)$ at an array of z's. Related to $E_1(z)$ are the sine and cosine integrals:

$$\begin{aligned} S_i(z) &= \int_0^z \frac{\sin u}{u}\,du && \text{(sine integral)} \\ C_i(z) &= \gamma + \ln z + \int_0^z \frac{\cos u - 1}{u}\,du && \text{(cosine integral)} \end{aligned} \tag{F.24}$$

where γ is the Euler constant $\gamma = 0.5772156649\dots$. A related cosine integral is:

$$C_{\text{in}}(z)=\int_0^z\frac{1-\cos u}{u}\,du = \gamma + \ln z - C_i(z) \tag{F.25}$$

For $z\ge 0$, the sine and cosine integrals are related to $E_1(z)$ by [1267]:

$$\begin{aligned} S_i(z) &= \frac{E_1(jz)-E_1(-jz)}{2j} + \frac{\pi}{2} = \text{Im}\big[E_1(jz)\big] + \frac{\pi}{2} \\ C_i(z) &= -\frac{E_1(jz)+E_1(-jz)}{2} = -\text{Re}\big[E_1(jz)\big] \end{aligned} \tag{F.26}$$

while for $z\le 0$, we have $S_i(z)=-S_i(-z)$ and $C_i(z)=C_i(-z)+j\pi$. Conversely, we have for $z>0$:

$$E_1(jz) = -C_i(z)+j\big(S_i(z)-\frac{\pi}{2}\big) = -\gamma - \ln(z) + C_{\text{in}}(z) + j\big(S_i(z)-\frac{\pi}{2}\big) \tag{F.27}$$

The MATLAB functions `Si`, `Ci`, `Cin` evaluate the sine and cosine integrals at any vector of z's by using the relations (F.26) and the built-in function `expint`:

```
y = Si(z);          % sine integral, Eq. (F.24)
y = Ci(z);          % sine integral, Eq. (F.24)
y = Cin(z);         % sine integral, Eq. (F.25)
```

A related integral that appears in calculating mutual and self impedances is what may be called a "Green's function integral":

$$\text{Gi}(d,z_0,h,s)=\int_0^h \frac{e^{-jkR}}{R}\,e^{-jksz}\,dz\,,\quad R=\sqrt{d^2+(z-z_0)^2}\,,\quad s=\pm 1 \tag{F.28}$$

This integral can be reduced to the exponential integral by the change of variables:

$$v = jk\big(R + s(z-z_0)\big) \quad\Rightarrow\quad s\frac{dv}{v}=\frac{dz}{R}$$

which gives

$$\int_0^h \frac{e^{-jkR}}{R}\,e^{-jksz}\,dz = se^{-jksz_0}\int_{v_0}^{v_1}\frac{e^{-u}}{u}\,du\,,\quad \text{or,}$$

$$\boxed{\text{Gi}(d,z_0,h,s)=\int_0^h \frac{e^{-jkR}}{R}\,e^{-jksz}\,dz = se^{-jksz_0}\big[E_1(ju_0)-E_1(ju_1)\big]} \tag{F.29}$$

where

$$v_0 = ju_0\,,\quad u_0 = k\left[\sqrt{d^2+z_0^2}-sz_0\right]$$

$$v_1 = ju_1\,,\quad u_1 = k\left[\sqrt{d^2+(h-z_0)^2}+s(h-z_0)\right]$$

The function `Gi` evaluates Eq. (F.29), where z_0, s, and the resulting integral J, can be vectors of the same dimension. Its usage is:

```
J = Gi(d,z0,h,s);        % Green's function integral, Eq. (F.29)
```

Another integral that appears commonly in antenna work is:

$$\int_0^{\pi}\frac{\cos(\alpha\cos\theta)-\cos\alpha}{\sin\theta}\,d\theta = S_i(2\alpha)\sin\alpha - C_{\text{in}}(2\alpha)\cos\alpha \tag{F.30}$$

Its proof is straightforward by first changing variables to $z=\cos\theta$, then using partial fraction expansion, and finally changing variables to $u=\alpha(1+z)$, and using the definitions (F.24) and (F.25):

$$\int_0^{\pi}\frac{\cos(\alpha\cos\theta)-\cos\alpha}{\sin\theta}\,d\theta = \int_{-1}^{1}\frac{\cos(\alpha z)-\cos\alpha}{1-z^2}\,dz$$

$$= \frac{1}{2}\int_{-1}^{1}\frac{\cos(\alpha z)-\cos\alpha}{1+z}\,dz + \frac{1}{2}\int_{-1}^{1}\frac{\cos(\alpha z)-\cos\alpha}{1-z}\,dz = \int_{-1}^{1}\frac{\cos(\alpha z)-\cos\alpha}{1+z}\,dz$$

$$= \int_0^{2\alpha}\frac{\cos(u-\alpha)-\cos\alpha}{u}\,du = \sin\alpha\int_0^{2\alpha}\frac{\sin u}{u}\,du - \cos\alpha\int_0^{2\alpha}\frac{1-\cos u}{u}\,du$$

G. Gauss-Legendre Quadrature

In many parts of this book it is necessary to perform numerical integration. Gauss-Legendre quadrature is one of the best integration methods, and we have implemented it with the MATLAB functions `quadr` and `quadrs`. Below, we give a brief description of the method.[†] The integral over an interval $[a,b]$ is approximated by a sum of the form:

$$\int_a^b f(x)\,dx \simeq \sum_{i=1}^{N} w_i f(x_i) \tag{G.1}$$

where w_i, x_i are appropriate weights and evaluation points (nodes). This can be written in the vectorial form:

$$\int_a^b f(x)\,dx \simeq \sum_{i=1}^{N} w_i f(x_i) = [w_1, w_2, \dots, w_N]\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N)\end{bmatrix} = \mathbf{w}^T f(\mathbf{x}) \tag{G.2}$$

The function `quadr` returns the column vectors of weights $\mathbf{w}$ and nodes $\mathbf{x}$, with usage:

```
[w,x] = quadr(a,b,N);      Gauss-Legendre quadrature
```

The function `quadrs` allows the splitting of the interval $[a,b]$ into subintervals, computes N weights and nodes in each subinterval, and concatenates them to form the overall weight and node vectors $\mathbf{w}, \mathbf{x}$:

```
[w,x] = quadrs(ab,N);      Gauss-Legendre quadrature over subintervals
```

where `ab` is an array of endpoints that define the subintervals, for example,

`ab` $= [a,b]$, single interval
`ab` $= [a,c,b]$, two subintervals, $[a,c]$ and $[c,b]$
`ab` $= [a,c,d,b]$, three subintervals, $[a,c]$, $[c,d]$, and $[d,b]$
`ab` $= a:c:b$, subintervals, $[a,\ a{+}c,\ a{+}2c,\ \dots,\ a{+}Mc]$, with $a + Mc = b$

As an example, consider the following function and its exact integral:

$$f(x) = e^x + \frac{1}{x}, \qquad J = \int_1^2 f(x)\,dx = e^2 - e^1 + \ln 2 = 5.36392145$$

This integral can be evaluated numerically by the MATLAB code:

```
N = 5;                     % number of weights and nodes
[w,x] = quadr(1,2,N);      % calculate weights and nodes for the interval [1,2]
f = exp(x) + 1./x;         % evaluate f(x) at the node vector
J = w'*f                   % approximate integral
```

This produces the exact value with a 4.23×10^{-7} percentage error. If the integration interval is split in two, say, $[1,1.5]$ and $[1.5,2]$, then the second line above can be replaced by

```
[w,x] = quadrs([1,1.5,2],N);     % or by, [w,x] = quadrs(1:0.5:2, N);
```

which has a percentage error of 1.28×10^{-9}. Next, we discuss the theoretical basis of the method.

The interval $[a,b]$ can be replaced by the standardized interval $[-1,1]$ with the transformation from $a \le x \le b$ to $-1 \le z \le 1$:

$$x = \left(\frac{b-a}{2}\right) z + \left(\frac{b+a}{2}\right) \tag{G.3}$$

If w_i and z_i are the weights and nodes with respect to the interval $[-1,1]$, then those with respect to $[a,b]$ can be constructed simply as follows, for $i = 1, 2, \dots, N$:

$$\begin{aligned} x_i &= \left(\frac{b-a}{2}\right) z_i + \left(\frac{b+a}{2}\right) \\ w_i^x &= \left(\frac{b-a}{2}\right) w_i \end{aligned} \tag{G.4}$$

where the scaling of the weights follows from the scaling of the differentials $dx = dz(b-a)/2$, so the value of the integral (G.1) is preserved by the transformation.

Gauss-Legendre quadrature is nicely tied with the theory of orthogonal polynomials over the interval $[-1,1]$, which are the Legendre polynomials. For N-point quadrature, the nodes z_i, $i = 1, 2, \dots, N$ are the N roots of the Legendre polynomial $P_N(z)$, which all lie in the interval $[-1,1]$. The method is justified by the following theorem:

For any polynomial $P(z)$ of degree at most $2N-1$, the quadrature formula (G.1) is satisfied *exactly*, that is,

$$\int_{-1}^{1} P(z)\,dz = \sum_{i=1}^{N} w_i P(z_i) \tag{G.5}$$

provided that the z_i are the N roots of the Legendre polynomial $P_N(z)$.

The Legendre polynomials $P_n(z)$ are obtained via the process of Gram-Schmidt orthogonalization of the non-orthogonal monomial basis $\{1, z, z^2, \dots, z^n \dots\}$. Orthogonality is defined with respect to the following inner product over the interval $[-1,1]$:

$$(f,g) = \int_{-1}^{1} f(z)g(z)\,dz \tag{G.6}$$

The standard definition of the Legendre polynomials is:

$$P_n(z) = \frac{1}{2^n n!}\frac{d^n}{dz^n}\left[(z^2-1)^n\right], \quad n = 0, 1, 2, \dots \tag{G.7}$$

The first few of them are listed below:

$$\begin{aligned} P_0(z) &= 1 \\ P_1(z) &= z \\ P_2(z) &= (3/2)\left[z^2 - (1/3)\right] \\ P_3(z) &= (5/2)\left[z^3 - (3/5)z\right] \\ P_4(z) &= (35/8)\left[z^4 - (6/7)z^2 + (3/35)\right] \end{aligned} \tag{G.8}$$

[†] J. Stoer and R. Burlisch, *Introduction to Numerical Analysis*, Springer, NY, (1980); and, G. H. Golub and J. H. Welsch, "Calculation of Gauss Quadrature Rules," *Math. Comput.*, **23**, 221 (1969).

They are normalized such that $P_n(1)=1$ and are mutually orthogonal with respect to (G.6), but do not have unit norm:

$$(P_n,P_m)=\int_{-1}^{1}P_n(z)P_m(z)\,dz=\frac{2}{2n+1}\,\delta_{nm} \tag{G.9}$$

Moreover, they satisfy the three-term recurrence relation:

$$zP_n(z)=\left(\frac{n}{2n+1}\right)P_{n-1}(z)+\left(\frac{n+1}{2n+1}\right)P_{n+1}(z) \tag{G.10}$$

The Gram-Schmidt orthogonalization process of the monomial basis $f_n(z)=z^n$ is the following order-recursive construction:

$$\begin{aligned}&\text{initialize}\quad P_0(z)=f_0(z)=1\\&\text{for } n=1,2,3,\dots,\ \text{do}\\&\qquad P_n(z)=f_n(z)-\sum_{k=0}^{n-1}\frac{(f_n,P_k)}{(P_k,P_k)}P_k(z)\end{aligned}$$

A few steps of the construction will clarify it:

$$P_1(z)=f_1(z)-\frac{(f_1,P_0)}{(P_0,P_0)}P_0(z)=z$$

where $(f_1,P_0)=(z,1)=\int_{-1}^{1}z\,dz=0$. Then, construct P_2 by:

$$P_2(z)=f_2(z)-\frac{(f_2,P_0)}{(P_0,P_0)}P_0(z)-\frac{(f_2,P_1)}{(P_1,P_1)}P_1(z)$$

where now we have $(f_2,P_1)=(z^2,z)=\int_{-1}^{1}z^3\,dz=0$, and

$$(f_2,P_0)=(z^2,1)=\int_{-1}^{1}z^2\,dz=\frac{2}{3}\,,\quad (P_0,P_0)=(1,1)=\int_{-1}^{1}dz=2$$

Therefore,

$$P_2(z)=z^2-\frac{2/3}{2}=z^2-\frac{1}{3}$$

Then, normalize it such that $P_2(1)=1$, and so on. For our discussion, we are going to renormalize the Legendre polynomials to unit norm. Because of (G.9), this amounts to multiplying the standard $P_n(z)$ by the factor $\sqrt{(2n+1)/2}$. Thus, we re-define:

$$P_n(z)=\sqrt{\frac{2n+1}{2}}\,\frac{1}{2^n n!}\,\frac{d^n}{dz^n}\big[(z^2-1)^n\big]\,,\quad n=0,1,2,\dots \tag{G.11}$$

Thus, (G.9) becomes $(P_n,P_m)=\delta_{nm}$. In particular, we note that now

$$P_0(z)=\frac{1}{\sqrt{2}} \tag{G.12}$$

By introducing the same scaling factors into each term of the recurrence (G.10), we find that the renormalized $P_n(z)$ satisfy:

$$zP_n(z)=\alpha_n P_{n-1}(z)+\alpha_{n+1}P_{n+1}(z)\,,\qquad \alpha_n=\frac{n}{\sqrt{4n^2-1}} \tag{G.13}$$

This relationship can be assumed to be valid also at $n=0$, provided we define $P_{-1}(z)=0$. For each order n, the Gram-Schmidt procedure replaces the non-orthogonal monomial basis by the orthonormalized Legendre basis:

$$\{1,\,z,\,z^2,\,\dots,\,z^n\}\ \Leftrightarrow\ \{P_0(z),\,P_1(z),\,P_2(z),\,\dots,\,P_n(z)\}$$

Thus, any polynomial $Q(z)$ of degree n can be expanded uniquely in either basis:

$$Q(z)=\sum_{k=0}^{n}q_k z^k=\sum_{k=0}^{n}c_k P_k(z)$$

with the expansion coefficients calculated from $c_k=(Q,P_k)$. This also implies that if $Q(z)$ has order $n-1$ then, it will be orthogonal to $P_n(z)$.

Next, we turn to the proof of the basic Gauss-Legendre result (G.5). Given a polynomial $P(z)$ of order $2N-1$, we can expand it uniquely in the form:

$$P(z)=P_N(z)Q(z)+R(z) \tag{G.14}$$

where $Q(z)$ and $R(z)$ are the quotient and remainder of the division by the Legendre polynomial $P_N(z)$, and both will have order $N-1$. Then, the integral of $P(z)$ can be written in inner-product notation as follows:

$$\int_{-1}^{1}P(z)\,dz=(P,1)=(P_NQ+R,1)=(P_NQ,1)+(R,1)=(Q,P_N)+(R,1)$$

But $(Q,P_N)=0$ because $Q(z)$ has order $N-1$ and $P_N(z)$ is orthogonal to all such polynomials. Thus, the integral of $P(z)$ can be expressed only in terms of the integral of the remainder polynomial $R(z)$, which has order $N-1$:

$$\int_{-1}^{1}P(z)\,dz=(P,1)=(R,1)=\int_{-1}^{1}R(z)\,dz \tag{G.15}$$

The right-hand side of the integration rule (G.5) can also be expressed in terms of $R(z)$:

$$\sum_{i=1}^{N}w_iP(z_i)=\sum_{i=1}^{N}w_iP_N(z_i)Q(z_i)+\sum_{i=1}^{N}w_iR(z_i) \tag{G.16}$$

and, because we assumed that $P_N(z_i)=0$,

$$\sum_{i=1}^{N}w_iP(z_i)=\sum_{i=1}^{N}w_iR(z_i) \tag{G.17}$$

Thus, combining (G.15) and (G.17), we obtain the following condition, which is equivalent to Eq. (G.5),

$$\int_{-1}^{1}R(z)\,dz=\sum_{i=1}^{N}w_iR(z_i) \tag{G.18}$$

Because $R(z)$ is an arbitrary polynomial of degree $N-1$, and has only N coefficients, this condition can be satisfied with a common set of N weights w_i for all such $R(z)$. If we had not assumed initially that the z_i were the zeros of $P_N(z)$, and took them to be an arbitrary set of N distinct points in $[-1,1]$, then (G.18) would read as

$$\int_{-1}^{1} R(z)\,dz = \sum_{i=1}^{N} w_i P_N(z_i)\,Q(z_i) + \sum_{i=1}^{N} w_i R(z_i)$$

In order for this to be satisfied for all $R(z)$ and all $Q(z)$, then (G.18) must still be satisfied by setting $Q(z)=0$, which fixes the weights w_i. Therefore, the first term in the right-hand side must be zero for all polynomials $Q(z)$ of degree $N-1$, and one can show that his implies that $P_N(z_i)=0$, that is, the z_i must be the zeros of $P_N(z)$.

Condition (G.18) can be used to determine the weights by expanding $R(z)$ into either the monomial basis or the Legendre basis, that is, because $R(z)$ has degree $N-1$:

$$R(z) = \sum_{k=0}^{N-1} r_k z^k = \sum_{k=0}^{N-1} c_k P_k(z) \tag{G.19}$$

Inserting, for example, the monomial basis into (G.18) and matching the coefficients of r_k on either side, we obtain the system of N equations for the weights:

$$\sum_{i=1}^{N} z_i^k w_i = \int_{-1}^{1} z^k\,dz = \frac{1+(-1)^k}{k+1}, \quad k = 0,1,\dots,N-1 \tag{G.20}$$

Defining the matrix $F_{ki} = z_i^k$ and the vector $u_k = [1+(-1)^k]/(k+1)$, we may write (G.20) in the compact matrix form:

$$F\mathbf{w} = \mathbf{u} \quad\Rightarrow\quad \mathbf{w} = F^{-1}\mathbf{u} \tag{G.21}$$

Alternatively, we may use the Legendre basis, which is more elegant. The left hand side of (G.18) will receive contribution only from the $k=0$ term because P_0 is orthogonal to all the succeeding P_k. Indeed, using the definition (G.12), we have:

$$\int_{-1}^{1} R(z)\,dz = (R,1) = \sqrt{2}\,(R,P_0) = \sqrt{2}\sum_{k=0}^{N-1} c_k (P_K,P_0) = \sqrt{2}\sum_{k=0}^{N-1} c_k\,\delta_{k0} = \sqrt{2}\,c_0$$

The right-hand side of (G.18) may be written as follows. Defining the $N{\times}N$ matrix $P_{ki} = P_k(z_i)$, $i = 1,2\dots,N$, and $k = 0,1,\dots,N-1$, and the row vector $\mathbf{c}^T = [c_0,c_1,\dots,c_{N-1}]$ of expansion coefficients, we have,

$$\sum_{i=1}^{N} w_i R(z_i) = \sum_{k=0}^{N-1}\sum_{i=1}^{N} c_k P_k(z_i)\,w_i = \mathbf{c}^T P\,\mathbf{w}$$

Thus, (G.18) now reads, where $\mathbf{u}_0 = [1,0,0,\dots,0]^T$:

$$\mathbf{c}^T P\,\mathbf{w} = \sqrt{2}\,c_0 = \sqrt{2}\,\mathbf{c}^T\mathbf{u}_0$$

Because the vector $\mathbf{c}$ is arbitrary, we must have the condition:

$$P\,\mathbf{w} = \sqrt{2}\,\mathbf{u}_0 \quad\Rightarrow\quad \mathbf{w} = \sqrt{2}\,P^{-1}\mathbf{u}_0 \tag{G.22}$$

The matrix P has some rather interesting properties. First, it has mutually orthogonal columns. Second, these columns are the eigenvectors of a Hermitian tridiagonal matrix whose eigenvalues are the zeros z_i. Thus, the problem of finding both z_i and w_i is reduced to an eigenvalue problem.

These eigenvalue properties follow from the recursion (G.13) of the normalized Legendre polynomials. For $n = 0,1,2,3$, the recursion reads explicitly:

$$\begin{aligned}
zP_0(z) &= \alpha_1 P_1(z)\\
zP_1(z) &= \alpha_1 P_0(z) + \alpha_2 P_2(z)\\
zP_2(z) &= \alpha_2 P_1(z) + \alpha_3 P_3(z)\\
zP_3(z) &= \alpha_3 P_2(z) + \alpha_4 P_4(z)
\end{aligned}$$

which can be written in matrix form:

$$z\begin{bmatrix} P_0(z)\\ P_1(z)\\ P_2(z)\\ P_3(z)\end{bmatrix} = \begin{bmatrix} 0 & \alpha_1 & 0 & 0\\ \alpha_1 & 0 & \alpha_2 & 0\\ 0 & \alpha_2 & 0 & \alpha_3\\ 0 & 0 & \alpha_3 & 0\end{bmatrix}\begin{bmatrix} P_0(z)\\ P_1(z)\\ P_2(z)\\ P_3(z)\end{bmatrix} + \begin{bmatrix} 0\\ 0\\ 0\\ \alpha_4 P_4(z)\end{bmatrix}$$

and more generally,

$$z\begin{bmatrix} P_0(z)\\ P_1(z)\\ P_2(z)\\ \vdots\\ P_{N-2}(z)\\ P_{N-1}(z)\end{bmatrix} = \begin{bmatrix} 0 & \alpha_1 & 0 & 0 & \cdots & 0\\ \alpha_1 & 0 & \alpha_2 & 0 & \cdots & 0\\ 0 & \alpha_2 & 0 & \alpha_3 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & \alpha_{N-2} & 0 & \alpha_{N-1}\\ 0 & \cdots & 0 & 0 & \alpha_{N-1} & 0\end{bmatrix}\begin{bmatrix} P_0(z)\\ P_1(z)\\ P_2(z)\\ \vdots\\ P_{N-2}(z)\\ P_{N-1}(z)\end{bmatrix} + \begin{bmatrix} 0\\ 0\\ 0\\ \vdots\\ 0\\ \alpha_N P_N(z)\end{bmatrix}$$

Now, if z is replaced by the ith zero z_i of $P_N(z)$, the last column will vanish and we obtain the eigenvalue equation:

$$\begin{bmatrix} 0 & \alpha_1 & 0 & 0 & \cdots & 0\\ \alpha_1 & 0 & \alpha_2 & 0 & \cdots & 0\\ 0 & \alpha_2 & 0 & \alpha_3 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & \alpha_{N-2} & 0 & \alpha_{N-1}\\ 0 & \cdots & 0 & 0 & \alpha_{N-1} & 0\end{bmatrix}\begin{bmatrix} P_0(z_i)\\ P_1(z_i)\\ P_2(z_i)\\ \vdots\\ P_{N-2}(z_i)\\ P_{N-1}(z_i)\end{bmatrix} = z_i\begin{bmatrix} P_0(z_i)\\ P_1(z_i)\\ P_2(z_i)\\ \vdots\\ P_{N-2}(z_i)\\ P_{N-1}(z_i)\end{bmatrix} \tag{G.23}$$

Denoting the above tridiagonal matrix by A and the column of $P_k(z_i)$'s by $\mathbf{p}_i$, we may write compactly:

$$A\,\mathbf{p}_i = z_i\,\mathbf{p}_i\,, \quad i = 1,2,\dots,N \tag{G.24}$$

Thus, the eigenvalues of A are the zeros z_i and the corresponding eigenvectors are the columns $\mathbf{p}_i$ of the matrix P that we introduced in (G.22). Because the zeros z_i are distinct and A is a Hermitian matrix, its eigenvectors will be mutually orthogonal:

$$\mathbf{p}_i^T\mathbf{p}_j = d_i^2\delta_{ij} \tag{G.25}$$

where $d_i = \|\mathbf{p}_i\|$ are the norms of the vectors $\mathbf{p}_i$. It follows that the orthonormalized eigenvectors of A will be $\boldsymbol{v}_i = \mathbf{p}_i/d_i$, and the orthogonal matrix of eigenvectors having the $\boldsymbol{v}_i$ as columns will be $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N]$, or, expressed in terms of the matrix P and the diagonal matrix $D = \text{diag}\{d_1, d_2, \dots, d_N\}$:

$$V = PD^{-1} \tag{G.26}$$

Replacing P in (G.22) by $P = VD$ and using the orthogonality $V^TV = I$ of the eigenvector matrix, or $V^{-1} = V^T$, we obtain the solution:

$$\mathbf{w} = \sqrt{2}\,D^{-1}V^T\mathbf{u}_0 \quad\Rightarrow\quad w_i = \sqrt{2}\,d_i^{-1}(\boldsymbol{v}_i^T\mathbf{u}_0) \tag{G.27}$$

The matrix D can itself be expressed in terms of V by noting that the top entry of $\mathbf{p}_i$ is $P_0(z_i) = 1/\sqrt{2}$, and therefore, it follows from $\boldsymbol{v}_i = \mathbf{p}_i/d_i$ that the top entry of $\boldsymbol{v}_i$ will be $\boldsymbol{v}_i^T\mathbf{u}_0 = 1/(\sqrt{2}d_i)$, or, $d_i^{-1} = \sqrt{2}(\boldsymbol{v}_i^T\mathbf{u}_0)$. It finally follows from Eq. (G.27) that

$$w_i = d_i^{-2} = 2(\boldsymbol{v}_i^T\mathbf{u}_0)^2 \tag{G.28}$$

In MATLAB language, $\boldsymbol{v}_i^T\mathbf{u}_0 = V(1,i)$, that is, the first row of V. Because the eigenvectors of the Hermitian matrix A are real-valued and unique up to a sign, Eq. (G.28) allows the unique determination of the weights from the eigenvector matrix V.

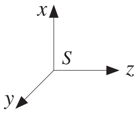
The above discussion leads to two possible implementations of the MATLAB function `quadr`. In the first, we obtain the coefficients of the Legendre polynomial $P_N(z)$, find its zeros using the built-in function `root`, and then solve the linear equation (G.21) for the weights. The second approach, implemented by the function `quadr2` and the related function `quadrs2`, determines z_i, w_i from the eigenvalue problem of the matrix A.

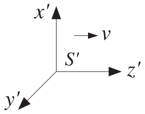
H. Lorentz Transformations

According to Einstein's special theory of relativity [435], Lorentz transformations describe the transformation between the space-time coordinates of two coordinate systems moving relative to each other at constant velocity. Maxwell's equations remain invariant under Lorentz transformations. This is demonstrated below.

Let the two coordinate frames be S and S'. By convention, we may think of S as the "fixed" laboratory frame with respect to which the frame S' is moving at a constant velocity $\boldsymbol{v}$. For example, if $\boldsymbol{v}$ is in the z-direction, the space-time coordinates $\{t, x, y, z\}$ of S are related to the coordinates $\{t', x', y', z'\}$ of S' by the Lorentz transformation:

$$\boxed{\begin{aligned} t' &= \gamma\Big(t - \frac{v}{c^2}z\Big)\\ z' &= \gamma(z - vt)\\ x' &= x\\ y' &= y \end{aligned}}\,, \quad \text{where} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$





where c is the speed of light in vacuum. Defining the scaled quantities $\tau = ct$ and $\beta = v/c$, the above transformation and its inverse, obtained by replacing β by $-\beta$, may be written as follows:

$$\boxed{\begin{aligned} \tau' &= \gamma(\tau - \beta z)\\ z' &= \gamma(z - \beta\tau)\\ x' &= x\\ y' &= y \end{aligned}} \quad\Leftrightarrow\quad \boxed{\begin{aligned} \tau &= \gamma(\tau' + \beta z')\\ z &= \gamma(z' + \beta\tau')\\ x &= x'\\ y &= y' \end{aligned}} \tag{H.1}$$

These transformations are also referred to as *Lorentz boosts* to indicate the fact that one frame is boosted to move relative to the other. Interchanging the roles of z and x, or z and y, one obtains the Lorentz transformations for motion along the x or y directions, respectively. Eqs. (H.1) may be expressed more compactly in matrix form:

$$\boxed{\mathsf{x}' = L\mathsf{x}}\,, \quad \text{where} \quad \mathsf{x} = \begin{bmatrix} \tau\\ x\\ y\\ z \end{bmatrix}, \quad \mathsf{x}' = \begin{bmatrix} \tau'\\ x'\\ y'\\ z' \end{bmatrix}, \quad L = \begin{bmatrix} \gamma & 0 & 0 & -\gamma\beta\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ -\gamma\beta & 0 & 0 & \gamma \end{bmatrix} \tag{H.2}$$

Such transformations leave the quadratic form $(c^2t^2 - x^2 - y^2 - z^2)$ invariant, that is,

$$c^2t'^2 - x'^2 - y'^2 - z'^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{H.3}$$

Introducing the diagonal metric matrix $G = \text{diag}(1, -1, -1, -1)$, we may write the quadratic form as follows, where x^T denotes the transposed vector, that is, the row vector $\mathsf{x}^T = [\tau, x, y, z]$:

$$\mathsf{x}^T G\mathsf{x} = \tau^2 - x^2 - y^2 - z^2 = c^2t^2 - x^2 - y^2 - z^2 \tag{H.4}$$

More generally, a Lorentz transformation is defined as any linear transformation $\mathsf{x}' = L\mathsf{x}$ that leaves the quadratic form $\mathsf{x}^T G\mathsf{x}$ invariant. The invariance condition requires that: $\mathsf{x}'^T G\mathsf{x}' = \mathsf{x}^T L^T G L\mathsf{x} = \mathsf{x}^T G\mathsf{x}$, or

$$\boxed{L^T G L = G} \tag{H.5}$$

In addition to the Lorentz boosts of Eq. (H.1), the more general transformations satisfying (H.5) include rotations of the three spatial coordinates, as well as time or space reflections. For example, a rotation has the form:

$$L = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0\\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right]$$

where R is a 3×3 orthogonal rotation matrix, that is, $R^TR = I$, where I is the 3×3 identity matrix. The most general Lorentz boost corresponding to arbitrary velocity $\boldsymbol{v} = [v_x, v_y, v_z]^T$ is given by:

$$L = \left[\begin{array}{c|c} \gamma & -\gamma\boldsymbol{\beta}^T\\ \hline -\gamma\boldsymbol{\beta} & I + \dfrac{\gamma^2}{\gamma+1}\boldsymbol{\beta}\boldsymbol{\beta}^T \end{array}\right], \quad \text{where} \quad \boldsymbol{\beta} = \frac{\boldsymbol{v}}{c}, \quad \gamma = \frac{1}{\sqrt{1 - \boldsymbol{\beta}^T\boldsymbol{\beta}}} \tag{H.6}$$

When $\boldsymbol{v} = [0,0,v]^T$, or $\boldsymbol{\beta} = [0,0,\beta]^T$, Eq. (H.6) reduces to (H.1). Defining $\beta = |\boldsymbol{\beta}| = \sqrt{\boldsymbol{\beta}^T\boldsymbol{\beta}}$ and the unit vector $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}/\beta$, and using the relationship $\gamma^2\beta^2 = \gamma^2 - 1$, it can be verified that the spatial part of the matrix L can be written in the form:

$$I + \frac{\gamma^2}{\gamma+1}\boldsymbol{\beta}\boldsymbol{\beta}^T = I + (\gamma-1)\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T \tag{H.7}$$

The set of matrices L satisfying Eq. (H.5) forms a group called the *Lorentz group*. In particular, the z-directed boosts of Eq. (H.2) form a commutative subgroup. Denoting these boosts by $L(\beta)$, the application of two successive boosts by velocity factors $\beta_1 = v_1/c$ and $\beta_2 = v_2/c$ leads to the combined boost $L(\beta) = L(\beta_1)L(\beta_2)$, where:

$$\boxed{\beta = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2}} \quad\Leftrightarrow\quad \boxed{v = \frac{v_1+v_2}{1+v_1v_2/c^2}} \tag{H.8}$$

with $\beta = v/c$. Eq. (H.8) is Einstein's relativistic velocity addition theorem. The same group property implies also that $L^{-1}(\beta) = L(-\beta)$. The proof of Eq. (H.8) follows from the following condition, where $\gamma_1 = 1/\sqrt{1-\beta_1^2}$ and $\gamma_2 = 1/\sqrt{1-\beta_2^2}$:

$$\begin{bmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & 0 & -\gamma_1\beta_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_1\beta_1 & 0 & 0 & \gamma_1 \end{bmatrix}\begin{bmatrix} \gamma_2 & 0 & 0 & -\gamma_2\beta_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_2\beta_2 & 0 & 0 & \gamma_2 \end{bmatrix}$$

A *four-vector* is a four-dimensional vector that transforms like the vector x under Lorentz transformations, that is, its components with respect to the two moving frames S and S' are related by:

$$\boxed{a' = La}\,, \quad \text{where} \quad a = \begin{bmatrix} a_0 \\ a_x \\ a_y \\ a_z \end{bmatrix}, \quad a' = \begin{bmatrix} a_0' \\ a_x' \\ a_y' \\ a_z' \end{bmatrix} \tag{H.9}$$

For example, under the z-directed boost of Eq. (H.1), the four-vector a will transform as:

$$\boxed{\begin{aligned} a_0' &= \gamma(a_0 - \beta a_z) \\ a_z' &= \gamma(a_z - \beta a_0) \\ a_x' &= a_x \\ a_y' &= a_y \end{aligned}} \quad\Leftrightarrow\quad \boxed{\begin{aligned} a_0 &= \gamma(a_0' + \beta a_z') \\ a_z &= \gamma(a_z' + \beta a_0') \\ a_x &= a_x' \\ a_y &= a_y' \end{aligned}} \tag{H.10}$$

Four-vectors transforming according to Eq. (H.9) are referred to as *contravariant*. Under the general Lorentz boost of Eq. (H.6), the spatial components of a that are *transverse* to the direction of the velocity vector $\boldsymbol{v}$ remain *unchanged*, whereas the *parallel* component transforms as in Eq. (H.10), that is, the most general Lorentz boost transformation for a four-vector takes the form:

$$\boxed{\begin{aligned} a_0' &= \gamma(a_0 - \beta a_\parallel) \\ a_\parallel' &= \gamma(a_\parallel - \beta a_0) \\ \boldsymbol{a}_\perp' &= \boldsymbol{a}_\perp \end{aligned}} \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}\,, \quad \beta = |\boldsymbol{\beta}|\,, \quad \boldsymbol{\beta} = \frac{\boldsymbol{v}}{c} \tag{H.11}$$

where $a_\parallel = \hat{\boldsymbol{\beta}}^T\boldsymbol{a}$ and $\boldsymbol{a} = [a_x, a_y, a_z]^T$ is the spatial part of a. Then,

$$\boldsymbol{a}_\parallel = \hat{\boldsymbol{\beta}}a_\parallel = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^T\boldsymbol{a}) \quad \text{and} \quad \boldsymbol{a}_\perp = \boldsymbol{a} - \boldsymbol{a}_\parallel = \boldsymbol{a} - \hat{\boldsymbol{\beta}}a_\parallel$$

Setting $\boldsymbol{\beta} = \beta\hat{\boldsymbol{\beta}}$ and using Eq. (H.7), the Lorentz transformation (H.6) gives:

$$\begin{bmatrix} a_0' \\ \boldsymbol{a}' \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma\beta\hat{\boldsymbol{\beta}}^T \\ -\gamma\beta\hat{\boldsymbol{\beta}} & I + (\gamma-1)\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T \end{bmatrix}\begin{bmatrix} a_0 \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \gamma(a_0 - \beta a_\parallel) \\ \boldsymbol{a} - \hat{\boldsymbol{\beta}}a_\parallel + \hat{\boldsymbol{\beta}}\gamma(a_\parallel - \beta a_0) \end{bmatrix}$$

from which Eq. (H.11) follows.

For any two four-vectors a, b, the quadratic form a^TGb remains invariant under Lorentz transformations, that is, $a'^TGb' = a^TGb$, or,

$$a_0'b_0' - \boldsymbol{a}'\cdot\boldsymbol{b}' = a_0b_0 - \boldsymbol{a}\cdot\boldsymbol{b}\,, \quad \text{where} \quad a = \begin{bmatrix} a_0 \\ \boldsymbol{a} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ \boldsymbol{b} \end{bmatrix} \tag{H.12}$$

Some examples of four-vectors are given in the following table:

four-vector	a_0	a_x	a_y	a_z
time and space	ct	x	y	z
frequency and wavenumber	ω/c	k_x	k_y	k_z
energy and momentum	E/c	p_x	p_y	p_z
charge and current densities	$c\rho$	J_x	J_y	J_z
scalar and vector potentials	φ	cA_x	cA_y	cA_z

(H.13)

For example, under the z-directed boost of Eq. (H.1), the frequency-wavenumber transformation will be as follows:

$$\boxed{\begin{aligned} \omega' &= \gamma(\omega - \beta ck_z) \\ k_z' &= \gamma\Big(k_z - \frac{\beta}{c}\omega\Big) \\ k_x' &= k_x \\ k_y' &= k_y \end{aligned}} \quad\Leftrightarrow\quad \boxed{\begin{aligned} \omega &= \gamma(\omega' + \beta ck_z') \\ k_z &= \gamma\Big(k_z' + \frac{\beta}{c}\omega'\Big) \\ k_x &= k_x' \\ k_y &= k_y' \end{aligned}}\,, \quad \beta c = v\,, \quad \frac{\beta}{c} = \frac{v}{c^2} \tag{H.14}$$

where we rewrote the first equations in terms of ω instead of ω/c. The change in frequency due to motion is the basis of the Doppler effect. The invariance property (H.12) applied to the space-time and frequency-wavenumber four-vectors reads:

$$\omega't' - \boldsymbol{k}'\cdot\mathbf{r}' = \omega t - \boldsymbol{k}\cdot\mathbf{r} \tag{H.15}$$

This implies that a uniform plane wave remains a uniform plane wave in all reference frames moving at a constant velocity relative to each other. Similarly, the charge and current densities transform as follows:

$$\boxed{\begin{aligned} c\rho' &= \gamma(c\rho - \beta J_z) \\ J_z' &= \gamma(J_z - \beta c\rho) \\ J_x' &= J_x \\ J_y' &= J_y \end{aligned}} \quad\Leftrightarrow\quad \boxed{\begin{aligned} c\rho &= \gamma(c\rho' + \beta J_z') \\ J_z &= \gamma(J_z' + \beta c\rho') \\ J_x &= J_x' \\ J_y &= J_y' \end{aligned}} \tag{H.16}$$

Because Eq. (H.5) implies that $L^{-T} = GLG$, we are led to define four-vectors that transform according to L^{-T}. Such four-vectors are referred to as being *covariant*. Given any contravariant 4-vector a, we define its covariant version by $\bar{a} = Ga$. This operation simply reverses the sign of the spatial part of a:

$$\bar{a} = Ga = \begin{bmatrix} 1 & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} a_0 \\ \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} a_0 \\ -\boldsymbol{a} \end{bmatrix} \tag{H.17}$$

The vector $\bar{a}$ transforms as follows:

$$\bar{a}' = Ga' = GLa = (GLG)(Ga) = L^{-T}\bar{a} \tag{H.18}$$

where we used the property that $G^2 = I_4$, the 4×4 identity matrix. The most important covariant vector is the four-dimensional gradient:

$$\partial_{\mathsf{x}} = \begin{bmatrix} \partial_\tau \\ \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} = \begin{bmatrix} \partial_\tau \\ \boldsymbol{\nabla} \end{bmatrix} \tag{H.19}$$

Because $\mathsf{x}' = L\mathsf{x}$, it follows that $\partial_{\mathsf{x}'} = L^{-T}\partial_{\mathsf{x}}$. Indeed, we have component-wise:

$$\frac{\partial}{\partial \mathsf{x}_i} = \sum_j \frac{\partial \mathsf{x}'_j}{\partial \mathsf{x}_i}\frac{\partial}{\partial \mathsf{x}'_j} = \sum_j L_{ji}\frac{\partial}{\partial \mathsf{x}'_j} \quad \Rightarrow \quad \partial_{\mathsf{x}} = L^T \partial_{\mathsf{x}'} \quad \Rightarrow \quad \partial_{\mathsf{x}'} = L^{-T}\partial_{\mathsf{x}}$$

For the z-directed boost of Eq. (H.1), we have $L^{-T} = L^{-1}$, which gives:

$$\boxed{\begin{aligned} \partial_{\tau'} &= \gamma(\partial_\tau + \beta\partial_z) \\ \partial_{z'} &= \gamma(\partial_z + \beta\partial_\tau) \\ \partial_{x'} &= \partial_x \\ \partial_{y'} &= \partial_y \end{aligned}} \quad \Leftrightarrow \quad \boxed{\begin{aligned} \partial_\tau &= \gamma(\partial_{\tau'} - \beta\partial_{z'}) \\ \partial_z &= \gamma(\partial_{z'} - \beta\partial_{\tau'}) \\ \partial_x &= \partial_{x'} \\ \partial_y &= \partial_{y'} \end{aligned}} \tag{H.20}$$

The four-dimensional divergence of a four-vector is a Lorentz scalar. For example, denoting the current density four-vector by $J = [c\rho, J_x, J_y, J_z]^T$, the charge conservation law involves the four-dimensional divergence:

$$\partial_t\rho + \boldsymbol{\nabla}\cdot\boldsymbol{J} = [\partial_\tau, \partial_x, \partial_y, \partial_z] \begin{bmatrix} c\rho \\ J_x \\ J_y \\ J_z \end{bmatrix} = \partial_{\mathsf{x}}^T J \tag{H.21}$$

Under a Lorentz transformation, this remains invariant, and therefore, if it is zero in one frame it will remain zero in all frames. Using $\partial_{\mathsf{x}}^T = \partial_{\mathsf{x}'}^T L$, we have:

$$\partial_t\rho + \boldsymbol{\nabla}\cdot\boldsymbol{J} = \partial_{\mathsf{x}}^T J = \partial_{\mathsf{x}'}^T LJ = \partial_{\mathsf{x}'}J' = \partial_{t'}\rho' + \boldsymbol{\nabla}'\cdot\boldsymbol{J}' \tag{H.22}$$

Although many quantities in electromagnetism transform like four-vectors, such as the space-time or the frequency-wavenumber vectors, the actual electromagnetic fields do not. Rather, they transform like six-vectors or rank-2 antisymmetric tensors.

A rank-2 tensor is represented by a 4×4 matrix, say F. Its Lorentz transformation properties are the same as the transformation of the product of a column and a row four-vector, that is, F transforms like the quantity ab^T, where a, b are column four-vectors. This product transforms like $a'b'^T = L(ab^T)L^T$. Thus, a general second-rank tensor transforms as follows:

$$\boxed{F' = LFL^T} \tag{H.23}$$

An *antisymmetric* rank-2 tensor F defines, and is completely defined by, two three-dimensional vectors, say $\boldsymbol{a} = [a_x, a_y, a_z]^T$ and $\boldsymbol{b} = [b_x, b_y, b_z]^T$. Its matrix form is:

$$F = \begin{bmatrix} 0 & -a_x & -a_y & -a_z \\ a_x & 0 & -b_z & b_y \\ a_y & b_z & 0 & -b_x \\ a_z & -b_y & b_x & 0 \end{bmatrix} \tag{H.24}$$

Given the tensor F, one may define its *covariant* version through $\bar{F} = GFG$, and its *dual*, denoted by $\tilde{F}$ and obtained by the replacements $\boldsymbol{a} \to \boldsymbol{b}$ and $\boldsymbol{b} \to -\boldsymbol{a}$, that is,

$$\bar{F} = \begin{bmatrix} 0 & a_x & a_y & a_z \\ -a_x & 0 & -b_z & b_y \\ -a_y & b_z & 0 & -b_x \\ -a_z & -b_y & b_x & 0 \end{bmatrix}, \quad \tilde{F} = \begin{bmatrix} 0 & -b_x & -b_y & -b_z \\ b_x & 0 & a_z & -a_y \\ b_y & -a_z & 0 & a_x \\ b_z & a_y & -a_x & 0 \end{bmatrix} \tag{H.25}$$

Thus, $\bar{F}$ corresponds to the pair $(-\boldsymbol{a}, \boldsymbol{b})$, and $\tilde{F}$ to $(\boldsymbol{b}, -\boldsymbol{a})$. Their Lorentz transformation properties are:

$$\bar{F}' = L^{-T}\bar{F}L^{-1}, \quad \tilde{F}' = L\tilde{F}L^T \tag{H.26}$$

Thus, the dual $\tilde{F}$ transforms like F itself. For the z-directed boost of Eq. (H.1), it follows from (H.23) that the two vectors $\boldsymbol{a}, \boldsymbol{b}$ transform as follows:

$$\boxed{\begin{aligned} a'_x &= \gamma(a_x - \beta b_y) & b'_x &= \gamma(b_x + \beta a_y) \\ a'_y &= \gamma(a_y + \beta b_x) & b'_y &= \gamma(b_y - \beta a_x) \\ a'_z &= a_z & b'_z &= b_z \end{aligned}} \tag{H.27}$$

These are obtained by equating the expressions:

$$\begin{bmatrix} 0 & -a'_x & -a'_y & -a'_z \\ a'_x & 0 & -b'_z & b'_y \\ a'_y & b'_z & 0 & -b'_x \\ a'_z & -b'_y & b'_x & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} 0 & -a_x & -a_y & -a_z \\ a_x & 0 & -b_z & b_y \\ a_y & b_z & 0 & -b_x \\ a_z & -b_y & b_x & 0 \end{bmatrix} \begin{bmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{bmatrix}$$

More generally, under the boost transformation (H.6), it can be verified that the components of $\boldsymbol{a}, \boldsymbol{b}$ parallel and perpendicular to $\boldsymbol{v}$ transform as follows:

$$\boxed{\begin{aligned} \boldsymbol{a}'_\perp &= \gamma(\boldsymbol{a}_\perp + \boldsymbol{\beta}\times\boldsymbol{b}_\perp)\\ \boldsymbol{b}'_\perp &= \gamma(\boldsymbol{b}_\perp - \boldsymbol{\beta}\times\boldsymbol{a}_\perp)\\ \boldsymbol{a}'_\parallel &= \boldsymbol{a}_\parallel\\ \boldsymbol{b}'_\parallel &= \boldsymbol{b}_\parallel \end{aligned}} \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}\,,\quad \beta = |\boldsymbol{\beta}|\,,\quad \boldsymbol{\beta} = \frac{\boldsymbol{v}}{c} \tag{H.28}$$

Thus, in contrast to Eq. (H.11) for a four-vector, the parallel components remain unchanged while the transverse components change. A pair of three-dimensional vectors $(\boldsymbol{a}, \boldsymbol{b})$ transforming like Eq. (H.28) is referred to as a *six-vector*.

It is evident also that Eqs. (H.28) remain invariant under the duality transformation $\boldsymbol{a} \to \boldsymbol{b}$ and $\boldsymbol{b} \to -\boldsymbol{a}$, which justifies Eq. (H.26). Some examples of $(\boldsymbol{a}, \boldsymbol{b})$ six-vector pairs defining an antisymmetric rank-2 tensor are as follows:

$$\begin{array}{cc} \boldsymbol{a} & \boldsymbol{b} \\ \hline \boldsymbol{E} & c\boldsymbol{B} \\ c\boldsymbol{D} & \boldsymbol{H} \\ c\boldsymbol{P} & -\boldsymbol{M} \end{array} \tag{H.29}$$

where $\boldsymbol{P}, \boldsymbol{M}$ are the polarization and magnetization densities defined through the relationships $\boldsymbol{D} = \epsilon_0\boldsymbol{E} + \boldsymbol{P}$ and $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$. Thus, the $(\boldsymbol{E}, \boldsymbol{B})$ and $(\boldsymbol{D}, \boldsymbol{H})$ fields have the following Lorentz transformation properties:

$$\boxed{\begin{aligned} \boldsymbol{E}'_\perp &= \gamma(\boldsymbol{E}_\perp + c\boldsymbol{\beta}\times\boldsymbol{B}_\perp)\\ \boldsymbol{B}'_\perp &= \gamma(\boldsymbol{B}_\perp - \frac{1}{c}\boldsymbol{\beta}\times\boldsymbol{E}_\perp)\\ \boldsymbol{E}'_\parallel &= \boldsymbol{E}_\parallel\\ \boldsymbol{B}'_\parallel &= \boldsymbol{B}_\parallel \end{aligned}} \qquad \boxed{\begin{aligned} \boldsymbol{H}'_\perp &= \gamma(\boldsymbol{H}_\perp - c\boldsymbol{\beta}\times\boldsymbol{D}_\perp)\\ \boldsymbol{D}'_\perp &= \gamma(\boldsymbol{D}_\perp + \frac{1}{c}\boldsymbol{\beta}\times\boldsymbol{H}_\perp)\\ \boldsymbol{H}'_\parallel &= \boldsymbol{H}_\parallel\\ \boldsymbol{D}'_\parallel &= \boldsymbol{D}_\parallel \end{aligned}} \tag{H.30}$$

where we may replace $c\boldsymbol{\beta} = \boldsymbol{v}$ and $\boldsymbol{\beta}/c = \boldsymbol{v}/c^2$. Note that the two groups of equations transform into each other under the usual duality transformations: $\boldsymbol{E} \to \boldsymbol{H}$, $\boldsymbol{H} \to -\boldsymbol{E}$, $\boldsymbol{D} \to \boldsymbol{B}$, $\boldsymbol{B} \to -\boldsymbol{D}$. For the z-directed boost of Eq. (H.1), we have from Eq. (H.30):

$$\boxed{\begin{aligned} E'_x &= \gamma(E_x - c\beta B_y)\\ E'_y &= \gamma(E_y + c\beta B_x)\\ B'_x &= \gamma(B_x + \frac{1}{c}\beta E_y)\\ B'_y &= \gamma(B_y - \frac{1}{c}\beta E_x)\\ E'_z &= E_z\\ B'_z &= B_z \end{aligned}} \qquad \boxed{\begin{aligned} H'_x &= \gamma(H_x + c\beta D_y)\\ H'_y &= \gamma(H_y - c\beta D_x)\\ D'_x &= \gamma(D_x - \frac{1}{c}\beta H_y)\\ D'_y &= \gamma(D_y + \frac{1}{c}\beta H_x)\\ H'_z &= H_z\\ D'_z &= D_z \end{aligned}} \tag{H.31}$$

Associated with a six-vector $(\boldsymbol{a}, \boldsymbol{b})$, there are two scalar invariants: the quantities $(\boldsymbol{a}\cdot\boldsymbol{b})$ and $(\boldsymbol{a}\cdot\boldsymbol{a} - \boldsymbol{b}\cdot\boldsymbol{b})$. Their invariance follows from Eq. (H.28). Thus, the scalars $(\boldsymbol{E}\cdot\boldsymbol{B})$, $(\boldsymbol{E}\cdot\boldsymbol{E} - c^2\boldsymbol{B}\cdot\boldsymbol{B})$, $(\boldsymbol{D}\cdot\boldsymbol{H})$, $(c^2\boldsymbol{D}\cdot\boldsymbol{D} - \boldsymbol{H}\cdot\boldsymbol{H})$ remain invariant under Lorentz transformations. In addition, it follows from (H.30) that the quantity $(\boldsymbol{E}\cdot\boldsymbol{D} - \boldsymbol{B}\cdot\boldsymbol{H})$ is invariant.

Given a six-vector $(\boldsymbol{a}, \boldsymbol{b})$ and its dual $(\boldsymbol{b}, -\boldsymbol{a})$, we may define the following four-dimensional "current" vectors that are dual to each other:

$$J = \begin{bmatrix} \nabla\cdot\boldsymbol{a} \\ \nabla\times\boldsymbol{b} - \partial_\tau\boldsymbol{a} \end{bmatrix}, \quad \tilde{J} = \begin{bmatrix} \nabla\cdot\boldsymbol{b} \\ -\nabla\times\boldsymbol{a} - \partial_\tau\boldsymbol{b} \end{bmatrix} \tag{H.32}$$

It can be shown that both J and $\tilde{J}$ transform as four-vectors under Lorentz transformations, that is, $J' = LJ$ and $\tilde{J}' = L\tilde{J}$, where $J', \tilde{J}'$ are defined with respect to the coordinates of the S' frame:

$$J' = \begin{bmatrix} \nabla'\cdot\boldsymbol{a}' \\ \nabla'\times\boldsymbol{b}' - \partial_{\tau'}\boldsymbol{a}' \end{bmatrix}, \quad \tilde{J}' = \begin{bmatrix} \nabla'\cdot\boldsymbol{b}' \\ -\nabla'\times\boldsymbol{a}' - \partial_{\tau'}\boldsymbol{b}' \end{bmatrix} \tag{H.33}$$

The calculation is straightforward but tedious. For example, for the z-directed boost (H.1), we may use Eqs. (H.20) and (H.27) and the identity $\gamma^2(1-\beta^2) = 1$ to show:

$$\begin{aligned} J'_x &= (\nabla'\times\boldsymbol{b}' - \partial_{\tau'}\boldsymbol{a}')_x = \partial_{y'}b'_z - \partial_{z'}b'_y - \partial_{\tau'}a'_x \\ &= \partial_y b_z - \gamma^2(\partial_z + \beta\partial_\tau)(b_y - \beta a_x) - \gamma^2(\partial_\tau + \beta\partial_z)(a_x - \beta b_y) \\ &= \partial_y b_z - \partial_z b_y - \partial_\tau a_x = (\nabla\times\boldsymbol{b} - \partial_\tau\boldsymbol{a})_x = J_x \end{aligned}$$

Similarly, we have:

$$\begin{aligned} J'_0 &= \nabla'\cdot\boldsymbol{a}' = \partial_{x'}a'_x + \partial_{y'}a'_y + \partial_{z'}a'_z \\ &= \gamma\partial_x(a_x - \beta b_y) + \gamma\partial_y(a_y + \beta b_x) + \gamma(\partial_z + \beta\partial_\tau)a_z \\ &= \gamma\big[(\partial_x a_x + \partial_y a_y + \partial_z a_z) - \beta(\partial_x b_y - \partial_y b_x - \partial_\tau a_z)\big] = \gamma(J_0 - \beta J_z) \end{aligned}$$

In this fashion, one can show that J and $\tilde{J}$ satisfy the Lorentz transformation equations (H.10) for a four-vector. To see the significance of this result, we rewrite Maxwell's equations, with magnetic charge and current densities $\rho_m, \boldsymbol{J}_m$ included, in the four-dimensional forms:

$$\begin{bmatrix} \nabla\cdot c\boldsymbol{D} \\ \nabla\times\boldsymbol{H} - \partial_\tau c\boldsymbol{D} \end{bmatrix} = \begin{bmatrix} c\rho \\ \boldsymbol{J} \end{bmatrix}, \quad \begin{bmatrix} \nabla\cdot c\boldsymbol{B} \\ -\nabla\times\boldsymbol{E} - \partial_\tau c\boldsymbol{B} \end{bmatrix} = \begin{bmatrix} c\rho_m \\ \boldsymbol{J}_m \end{bmatrix} \tag{H.34}$$

Thus, applying the above result to the six-vector $(c\boldsymbol{D}, \boldsymbol{H})$ and to the dual of $(\boldsymbol{E}, c\boldsymbol{B})$ and assuming that the electric and magnetic current densities transform like four-vectors, it follows that Maxwell's equations remain invariant under Lorentz transformations, that is, they retain their form in the moving system:

$$\begin{bmatrix} \nabla'\cdot c\boldsymbol{D}' \\ \nabla'\times\boldsymbol{H}' - \partial_{\tau'}c\boldsymbol{D}' \end{bmatrix} = \begin{bmatrix} c\rho' \\ \boldsymbol{J}' \end{bmatrix}, \quad \begin{bmatrix} \nabla'\cdot c\boldsymbol{B}' \\ -\nabla'\times\boldsymbol{E}' - \partial_{\tau'}c\boldsymbol{B}' \end{bmatrix} = \begin{bmatrix} c\rho'_m \\ \boldsymbol{J}'_m \end{bmatrix} \tag{H.35}$$

The Lorentz transformation properties of the electromagnetic fields allow one to solve problems involving moving media, such as the Doppler effect, reflection and transmission from moving boundaries, and so on. The main technique for solving such problems is to transform to the frame (here, S') in which the boundary is at rest, solve the reflection problem in that frame, and transform the results back to the laboratory frame by using the inverse of Eq. (H.30).

This procedure was discussed by Einstein in his 1905 paper on special relativity in connection to the Doppler effect from a moving mirror. To quote [435]: "All problems in the optics of moving bodies can be solved by the method here employed. What is essential is that the electric and magnetic force of the light which is influenced by a moving body, be transformed into a system of co-ordinates at rest relatively to the body. By this means all problems in the optics of moving bodies will be reduced to a series of problems in the optics of stationary bodies."

I. MATLAB Functions

The MATLAB functions are grouped by category. They are available from the web page: `www.ece.rutgers.edu/~orfanidi/ewa`.

Multilayer Dielectric Structures

```
brewster   - calculates Brewster and critical angles
fresnel    - Fresnel reflection coefficients for isotropic or birefringent media

n2r        - refractive indices to reflection coefficients of M-layer structure
r2n        - reflection coefficients to refractive indices of M-layer structure

multidiel  - reflection response of isotropic or birefringent multilayer structures
multidiel1 - simplified version of multidiel for isotropic layers
multidiel2 - reflection response of lossy isotropic multilayer dielectric structures
omniband   - bandwidth of omnidirectional mirrors and Brewster polarizers
omniband2  - bandwidth of birefringent multilayer mirrors

snel       - calculates refraction angles from Snel's law for birefringent media
```

Quarter-Wavelength Transformers

```
bkwrec  - order-decreasing backward layer recursion - from a,b to r
frwrec  - order-increasing forward layer recursion - from r to A,B

chebtr  - Chebyshev broadband reflectionless quarter-wave transformer
chebtr2 - Chebyshev broadband reflectionless quarter-wave transformer
chebtr3 - Chebyshev broadband reflectionless quarter-wave transformer
```

Dielectric Waveguides

```
dguide - TE modes in dielectric slab waveguide
dslab  - solves for the TE-mode cutoff wavenumbers in a dielectric slab
```

Transmission Lines

```
g2z       - reflection coefficient to impedance transformation
z2g       - impedance to reflection coefficient transformation
lmin      - find locations of voltage minima and maxima

mstripa   - microstrip analysis (calculates Z,eff from w/h)
mstripr   - microstrip synthesis with refinement (calculates w/h from Z)
mstrips   - microstrip synthesis (calculates w/h from Z)

multiline - reflection response of multi-segment transmission line

swr       - standing wave ratio
tsection  - T-section equivalent of a length-l transmission line segment

gprop     - reflection coefficient propagation
vprop     - wave impedance propagation
zprop     - wave impedance propagation
```

Impedance Matching

```
qwt1     - quarter wavelength transformer with series segment
qwt2     - quarter wavelength transformer with 1/8-wavelength shunt stub
qwt3     - quarter wavelength transformer with shunt stub of adjustable length

dualband - two-section dual-band Chebyshev impedance transformer
dualbw   - two-section dual-band transformer bandwidths

stub1    - single-stub matching
stub2    - double-stub matching
stub3    - triple-stub matching

onesect  - one-section impedance transformer
twosect  - two-section impedance transformer

pi2t     - Pi to T transformation
t2pi     - Pi to T transformation
lmatch   - L-section reactive conjugate matching network
pmatch   - Pi-section reactive conjugate matching network
```

S-Parameters

```
gin      - input reflection coefficient in terms of S-parameters
gout     - output reflection coefficient in terms of S-parameters
nfcirc   - constant noise figure circle
nfig     - noise figure of two-port
sgain    - transducer, available, and operating power gains of two-port
sgcirc   - stability and gain circles
smat     - S-parameters to S-matrix
smatch   - simultaneous conjugate match of a two-port
smith    - draw basic Smith chart
smithcir - add stability and constant gain circles on Smith chart
sparam   - stability parameters of two-port
circint  - circle intersection on Gamma-plane
circtan  - point of tangency between the two circles
```

Linear Antenna Functions

```
dipdir    - dipole directivity
dmax      - computes directivity and beam solid angle of g(th) gain
dipole    - gain of center-fed linear dipole of length L
traveling - gain of traveling-wave antenna of length L
vee       - gain of traveling-wave vee antenna
rhombic   - gain of traveling-wave rhombic antenna

king      - King’s 3-term sinusoidal approximation
kingeval  - evaluate King’s 3-term sinusoidal current approximation
kingfit   - fits a sampled current to King’s 2-term sinusoidal approximation
kingprime - converts King’s 3-term coefficients from unprimed to primed form

hbasis    - basis functions for Hallen equation
hdelta    - solve Hallen’s equation with delta-gap input
hfield    - solve Hallen’s equation with arbitrary incident E-field
hmat      - Hallen impedance matrix with method of moments and point-matching
hwrap     - wraps a Toeplitz impedance matrix to half its size
kernel    - thin-wire kernel computation for Hallen equation
pfield    - solve Pocklington’s equation with arbitrary incident E-field
pmat      - Pocklington impedance matrix with method of moments and point-matching

hcoupled  - solve Hallen’s equation for 2D array of non-identical parallel dipoles
hcoupled2 - solve Hallen’s equation for 2D array of identical parallel dipoles

gain2d    - normalized gain of 2D array of parallel dipoles with Hallen currents
gain2s    - normalized gain of 2D array of parallel dipoles with sinusoidal currents
imped     - mutual impedance between two parallel standing-wave dipoles
imped2    - mutual impedance between two parallel standing-wave dipoles
impedmat  - mutual impedance matrix of array of parallel dipole antennas
resonant  - calculates the length of a resonant dipole antenna
yagi      - simplified Yagi-Uda array design
```

Aperture Antenna Functions

```
diffint - generalized Fresnel diffraction integral
diffr   - knife-edge diffraction coefficient
dsinc   - the double-sinc function cos(pi*x)/(1-4*x^2)

fcs     - Fresnel integrals C(x) and S(x)
fcs2    - type-2 Fresnel integrals C2(x) and S2(x)

hband   - horn antenna 3-dB width
heff    - aperture efficiency of horn antenna
hgain   - horn antenna H-plane and E-plane gains
hopt    - optimum horn antenna design
hsigma  - optimum sigma parametes for horn antenna
```

Antenna Array Functions

```
gain1d    - normalized gain computation for 1D equally-spaced isotropic array
bwidth    - beamwidth mapping from psi-space to phi-space
binomial  - binomial array weights
dolph     - Dolph-Chebyshev array weights
dolph2    - Riblet-Pritchard version of Dolph-Chebyshev
dolph3    - DuHamel version of endfire Dolph-Chebyshev
```

```
multibeam - multibeam array design
prol      - prolate array design
prolmat   - prolate matrix
scan      - scan array with given scanning phase
sector    - sector beam array design
steer     - steer array towards given angle
taylor1n  - Taylor n-bar line source array design
taylor1p  - Taylor 1-parameter array design
taylorbw  - Taylor B-parameter and beamwidth
uniform   - uniform array weights
woodward  - Woodward-Lawson-Butler beams
ville     - Villeneuve array design

chebarray - Bresler’s Chebyshev array design method (written by P. Simon)
```

Gain Plotting Functions

```
abp     - polar gain plot in absolute units
abz     - azimuthal gain plot in absolute units
ab2p    - polar gain plot in absolute units - 2*pi angle range
abz2    - azimuthal gain plot in absolute units - 2pi angle range

dbp     - polar gain plot in dB
dbz     - azimuthal gain plot in dB
dbp2    - polar gain plot in dB - 2*pi angle range
dbz2    - azimuthal gain plot in dB - 2pi angle range

abadd   - add gain in absolute units
abadd2  - add gain in absolute units - 2pi angle range
dbadd   - add gain in dB
dbadd2  - add gain in dB - 2pi angle range
addbwp  - add 3-dB angle beamwidth in polar plots
addbwz  - add 3-dB angle beamwidth in azimuthal plots
addcirc - add grid circle in polar or azimuthal plots
addline - add grid ray line in azimuthal or polar plots
addray  - add ray in azimuthal or polar plots
```

Miscellaneous Utility Functions

```
ab       - dB to absolute power units
db       - absolute power to dB units

c2p      - complex number to phasor form
p2c      - phasor form to complex number

d2r      - degrees to radians
r2d      - radians to degrees

dtft     - DTFT of a signal x at a frequency vector w
I0       - modified Bessel function of 1st kind and 0th order
ellipse  - polarization ellipse parameters
etac     - eta and c
wavenum  - calculate wavenumber and characteristic impedance
poly2    - specialized version of poly with increased accuracy

quadr    - Gauss-Legendre quadrature weights and evaluation points
quadrs   - quadrature weights and evaluation points on subintervals
```

quadr2 - Gauss-Legendre quadrature weights and evaluation points
quadrs2 - quadrature weights and evaluation points on subintervals

Ci - cosine integral Ci(z)
Cin - cosine integral Cin(z)
Si - sine integral Si(z)
Gi - Green's function integral

sinhc - hyperbolic sinc function
asinhc - inverse hyperbolic sinc function
sqrte - evanescent SQRT for waves problems

flip - flip a column, a row, or both
blockmat - manipulate block matrices
upulse - generates trapezoidal, rectangular, triangular pulses, or a unit-step
ustep - unit-step or rising unit-step function

dnv - dn elliptic function at a vector of moduli
snv - sn elliptic function at a vector of moduli
ellipK - complete elliptic integral of first kind at a vector of moduli
ellipE - complete elliptic integral of second kind at a vector of moduli
landenv - Landen transformations of a vector of elliptic moduli

MATLAB Movies

grvmovie1 - pulse propagation with slow and negative group velocity ($v_g < 0$)
grvmovie2 - pulse propagation with slow and fast group velocity ($v_g > c$)
pulsemovie - step and pulse propagation on terminated transmission lines
pulse2movie - step propagation on two cascaded lines
RLCmovie - step getting reflected off a reactive termination
TDRmovie - fault location by time-domain reflectometry
xtalkmovie - crosstalk signals on coupled transmission lines
dipmovie - electric field pattern of radiating Hertzian dipole

References

Books

[1] S. A. Schelkunoff, *Electromagnetic Waves*, Van Nostrand, New York, 1943.

[2] S. A. Schelkunoff and H. T. Friis, *Antennas, Theory and Practice*, Wiley, New York, 1952.

[3] R. W. P. King, *The Theory of Linear Antennas*, Harvard Univ. Press, Cambridge, MA, 1956.

[4] R. W. P. King, R. B. Mack, and S. S. Sandler, *Arrays of Cylindrical Dipoles*, Cambridge Univ. Press, Cambridge, 1968.

[5] J. D. Kraus, *Antennas*, 2nd ed., McGraw-Hill, New York, 1988.

[6] C. A. Balanis, *Antenna Theory, Analysis and Design*, 2nd ed., Wiley, New York, 1996.

[7] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., Wiley, New York, 1998.

[8] R. S. Elliott, *Antenna Theory and Design*, Prentice Hall, Upper Saddle River, NJ, 1981.

[9] R. E. Collin and F. J. Zucker, eds., *Antenna Theory*, parts 1 and 2, McGraw-Hill, New York, 1969.

[10] A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, eds., *The Handbook of Antenna Design*, vols. 1 and 2, 2nd ed., Peter Peregrinus Ltd., London, 1986.

[11] R. C. Johnson, ed., *Antenna Engineering Handbook*, 3d ed., McGraw-Hill, New York, 1993.

[12] T. S. M. Maclean, *Principles of Antennas: Wire and Aperture*, Cambridge Univ. Press, Cambridge, 1986.

[13] J. R. Wait, *Introduction to Antennas and Propagation*, Peter Peregrinus, Ltd, London, 1986.

[14] T. A. Milligan, *Modern Antenna Design*, McGraw-Hill, New York, 1985.

[15] R. C. Hansen, *Microwave Scanning Antennas*, Academic Press, New York, vol. I, 1964, vols II and III, 1966.

[16] R. C. Hansen, *Phased Array Antennas*, Wiley, New York, 1998.

[17] E. Brookner, ed., *Practical Phased Array Antenna Systems*, Artech House, Boston, 1991.

[18] R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Norwood, MA, 1994.

[19] R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, New York, 1985.

[20] H. Mott, *Antennas for Radar and Communications: A Polarimetric Approach*, Wiley, New York, 1992.

[21] S. Silver, ed., *Microwave Antenna Theory and Design*, Peter Peregrinus, Ltd, London, 1984.

[22] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill, New York, 1961.

[23] G. S. Smith, *An Introduction to Classical Electromagnetic Radiation*, Cambridge Univ. Press, Cambridge, 1997.

[24] C. G. Someda, *Electromagnetic Waves*, Chapman and Hall, London, 1998.

[25] E. J. Rothwell and M. J. Cloud, *Electromagnetics*, CRC Press, Boca Raton, FL, 2001.

[26] L. B. Felsen and N, Marcuvitz, *Radiation and Scattering of Waves*, IEEE Press, New York, 1994.

[27] L. Tsang, J. A. Kong, and K-H. Ding, *Scattering of Electromagnetic Waves*, Wiley, New York, 2000.

[28] S. Drabowitch, A. Papiernik, H. Griffiths, and J. Encinas, *Modern Antennas*, Chapman & Hall, London, 1998.

[29] A. D. Olver, *Microwave and Optical Transmission*, Wiley, Chichester, England, 1992.

[30] J. A. Kong, *Electromagnetic Wave Theory*, 2nd ed., Wiley, New York, 1990.

[31] H. C. Chen, *Theory of Electromagnetic Waves*, McGraw Hill, New York, 1983.

[32] G. Tyras, *Radiation and Propagation of Electromagnetic Waves*, Academic Press, New York, 1969.

[33] D. H. Staelin, A. W. Morgenthaler, and J. A. Kong, *Electromagnetic Waves*, Prentice Hall, Upper Saddle River, NJ, 1994.

[34] E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1968.

[35] W. L. Weeks, *Antenna Engineering*, McGraw-Hill, New York, 1968.

[36] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3d ed., Wiley, New York, 1994.

[37] J. D. Kraus, *Electromagnetics*, 3d ed., McGraw-Hill, New York, 1984.

[38] D. J. Griffiths, *Introduction to Electrodynamics*, 3/e, Prentice Hall, Upper Saddle River, NJ, 1999.

[39] C. A. Balanis, *Advanced Engineering Electromagnetics*, Wiley, New York, 1989.

[40] J. D. Jackson, *Classical Electrodynamics*, 3d ed., Wiley, New York, 1998.

[41] M. A. Heald and J. B. Marion, *Classical Electromagnetic Radiation*, 3d ed., Saunders College Publishing, New York, 1995.

[42] A. A. Smith, *Radio Frequency Principles and Applications*, IEEE Press, Piscataway, NJ, 1998.

[43] R. D. Straw, ed., *The ARRL Antenna Book*, 18th ed., American Radio Relay League, Newington, CT, 1997.

[44] P. Danzer, ed., *The ARRL Handbook*, 75th ed., American Radio Relay League, Newington, CT, 1998.

[45] F. A. Benson and T. M. Benson, *Fields, Waves and Transmission Lines*, Chapman and Hall, London, 1991.

[46] A. F. Wickersham, *Microwave and Fiber Optics Communications*, Prentice Hall, Upper Saddle River, NJ, 1988.

[47] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1999.

[48] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice Hall, Upper Saddle River, NJ, 1996.

[49] D. K. Lynch and W. Livingston, *Color and Light in Nature*, Oxford Univ. Press, New York, 1995.

[50] M. Minnaert, *Light and Color in the Outdoors*, translated and revised by L. Seymour, Springer-Verlag, New York, 1993.

[51] A. Meinel and M. Meinel, *Sunsets, Twilights, and Evening Skies*, Cambridge University Press, Cambridge, 1983.

[52] G. P. Können, *Polarized Light in Nature*, Cambridge University Press, Cambridge, 1985.

[53] D. H. Towne, *Wave Phenomena*, Dover Publications, New York, 1988.

[54] W. C. Elmore and M. A. Heald, *Physics of Waves*, Dover Publications, New York, 1985.

[55] J. R. Peirce, *Almost All About Waves*, MIT Press, Cambridge, MA, 1974.

History, Reviews, Physical Constants, Units, and IEEE Standards

[56] H. Hertz, *Electric Waves*, Dover Publications, New York, 1962.

[57] M. R. Cohen and I. E. Drabkin, *A Source Book in Greek Science*, Harvard University Press, Cambridge, MA, 1969.

[58] A. M. Smith, "Ptolemy's Search for a Law of Refraction: A Case Study in the Classical Methodology of "Saving the Appearances" and its Limitations", *Arch. Hist. Exact Sci.*, **26**, 221 (1982).

[59] L. S. Taylor, "Gallery of Electromagnetic Personalities: A Vignette History of Electromagnetics," see web site Ref. [1301].

[60] L. S. Taylor, "Optics Highlights: An Anecdotal History of Optics from Aristophanes to Zernike," see web site Ref. [1302].

[61] J. Hecht, "Illuminating the Origin of Light Guiding," *Optics & Photonics News*, **10**, no.10, 26 (1999).

[62] H. Crew, "Thomas Young's Place in the History of the Wave Theory of Light," *J. Opt. Soc. Am.*, **20**, 3 (1930).

[63] E. Weber, "The Evolution of Scientific Electrical Engineering," *IEEE Antennas and Propagation Mag.*, **33**, 12, February 1991.

[64] J. J. Roche, "A Critical Study of the Vector Potential," in *Physicists Look Back, Studies in the History of Physics*, J. J. Roche, ed., Adam Hilger, Bristol, 1990.

[65] J. Van Bladel, "Lorentz or Lorenz?," *IEEE Antennas and Propagation Mag.*, **33**, 69, April 1991.

[66] J. D. Jackson and L. B. Okun, "Historical Roots of Gauge Invariance," *Rev. Mod. Phys.*, **73**, 663 (2001).

[67] R. Nevels and C-S. Shin, "Lorenz, Lorentz, and the Gauge," *IEEE Antennas and Propagation Mag.*, **43**, 70, April 2001.

[68] M. Pihl, "The Scientific Achievements of L. V. Lorenz," *Centaurus*, **17**, 83 (1972).

[69] Ludvig Lorenz and Nineteenth Century Optical Theory: The Work of a Great Danish Scientist," *Appl. Opt.*, **30**, 4688 (1991).

[70] O. Keller, "Optical Works of L. V. Lorenz," in *Progress in Optics*, vol.43, E. Wolf, ed., Elsevier, Amsterdam, 2002.

[71] R. W. P. King, "The Linear Antenna—Eighty Years of Progress," *Proc. IEEE*, **55**, 2 (1967).

[72] Chen-To Tai, "On the Presentation of Maxwell's Theory," *Proc. IEEE*, **60**, 936 (1972).

[73] J. A. Kong, "Theorems of Bianisotropic Media," *Proc. IEEE*, **60**, 1036 (1972).

[74] S. A. Schelkunoff, "Forty Years Ago: Maxwell's Theory Invades Engineering—and Grows with It," *IEEE Trans. Education*, **E-15**, 2 (1972).

[75] S. A. Schelkunoff, "On Teaching the Undergraduate Electromagnetic Theory," *IEEE Trans. Education*, **E-15**, 2 (1972).

[76] J. E. Brittain, "The Smith Chart," *IEEE Spectrum*, **29**, 65, Aug. 1992.

[77] G. McElroy, "Opening Lines: A Short History of Coaxial cable," *QST*, **85**, 62, Aug. 2001.

[78] V. J. Katz, "The History of Stokes' Theorem," *Math. Mag.*, **52**, 146 (1979).

[79] J. R. Jones, "A Comparison of Lightwave, Microwave, and Coaxial Transmission Technologies," *IEEE Trans. Microwave Theory Tech.*, **MTT-30**, 1512 (1982).

[80] D. K. Barton, "A Half Century of Radar," *IEEE Trans. Microwave Theory Tech.*, **MTT-32**, 1161 (1984).

[81] M. Kahrs, "50 Years of RF and Microwave Sampling," *IEEE Trans. Microwave Theory Tech.*, **MTT-51**, 1787 (2003).

[82] C. A. Balanis, "Antenna Theory: A Review," *Proc. IEEE*, **80**, 7 (1992).

[83] W. Stutzman, "Bibliography for Antennas," *IEEE Antennas and Propagation Mag.*, **32**, 54, August 1990.

[84] J. F. Mulligan, "The Influence of Hermann von Helmholtz on Heinrich Hertz's Contributions to Physics," *Am. J. Phys.*, **55**, 711 (1987).

[85] C-T. Tai and J. H. Bryant, "New Insights in Hertz's Theory of Electromagnetism," *Radio Science*, **29**, 685 (1994).

[86] R. Cecchini and G. Pelosi, "Diffraction: The First Recorded Observation," *IEEE Antennas and Propagation Mag.*, **32**, 27, April 1990.

[87] O. M. Bucci and G. Pelosi, "From Wave Theory to Ray Optics," *IEEE Antennas and Propagation Mag.*, **36**, 35, August 1994.

[88] A. A. Penzias, "Measurement of Cosmic Microwave Background Radiation," *IEEE Trans. Microwave Theory Tech.*, **MTT-16**, 608 (1968).

[89] E. R. Cohen, "The 1986 CODATA Recommended Values of the Fundamental Physical Constants," *J. Res. Natl. Bur. Stand.*, **92**, 85, (1987).

[90] C. W. Allen, *Astrophysical Quantities*, Athlone Press, University of London, London, 1976.

[91] "IEEE Standard Test Procedures for Antennas," IEEE Std 149-1965, *IEEE Trans. Antennas Propagat.*, **AP-13**, 437 (1965). Revised IEEE Std 149-1979.

[92] "IEEE Standard Definitions of Terms for Antennas," IEEE Std 145-1983, *IEEE Trans. Antennas Propagat.*, **AP-31**, pt.II, p.5, (1983). Revised IEEE Std 145-1993.

[93] "IRE Standards on Antennas and Waveguides: Definitions and Terms, 1953," *Proc. IRE*, **41**, 1721 (1953). Revised IEEE Std 146-1980.

[94] "IRE Standards on Methods of Measuring Noise in Linear Twoports, 1959," *Proc. IRE*, **48**, 60 (1960).

[95] IRE Subcommittee 7.9 on Noise,"Representation of Noise in Linear Twoports," *Proc. IRE*, **48**, 69 (1960).

[96] "IRE Standards on Electron Tubes: Definitions of Terms, 1962 (62 IRE 7. S2)," *Proc. IRE*, **51**, 434 (1963).

[97] IRE Subcommittee 7.9 on Noise,"Description of the Noise Performance of Amplifiers and Receiving Systems," *Proc. IEEE*, **51**, 436 (1963).

[98] C. K. S. Miller, W. C. Daywitt, and M. G. Arthur, "Noise Standards, Measurements, and Receiver Noise Definitions," *Proc. IEEE*, **55**, 865 (1967).

[99] A. R. Kerr, "Suggestions for Revised Definitions of Noise Quantities, Including Quantum Effects," *Proc. IEEE*, **47**, 325 (1999).

[100] E. S. Weibel, "Dimensionally Correct Transformations Between Different Systems of Units," *Am. J. Phys.*, **36**, 1130 (1968).

[101] B. Leroy, "How to Convert the Equations of Electromagnetism from Gaussian to SI Units in Less Than No Time," *Am. J. Phys.*, **53**, 589 (1985).

[102] E. A. Desloge, "Relation Between Equations in the International, Electrostatic, Electromagnetic, Gaussian, and Heaviside-Lorentz Systems," *Am. J. Phys.*, **62**, 602 (1994), and references therein on units.

[103] *NRL Plasma Formulary*, 2000 Revised Edition, see web site [1330].

[104] "IEEE Standard Radar Definitions," IEEE Std 686-1997.

[105] "IEEE Standard Letter Designations for Radar-Frequency Bands," IEEE Std 521®-2002.

Green's Functions

[106] J. Van Bladel, "Some Remarks on Green's Dyadic for Infinite Space," *IRE Trans. Antennas Propagat.*, **AP-9**, 563 (1961).

[107] K-M. Chen, "A Simple Physical Picture of Tensor Green's Function in Source Region," *Proc. IEEE*, **65**, 1202 (1977).

[108] A. D. Yaghjian, "Electric Dyadic Green's Functions in the Source Region," *Proc. IEEE*, **68**, 248 (1980). See also the comment by C. T. Tai, *ibid.*, **69**, 282 (1981).

[109] A. D. Yaghjian, "Maxwellian and Cavity Electromagnetic Fields Within Continuous Sources," *Am. J. Phys.*, **53**, 859 (1985).

[110] M. Silberstein, "Application of a Generalized Leibniz Rule for Calculating Electromagnetic Fields Within Continuous Source Regions," *Radio Sci.*, **26**, 183 (1991).

[111] C. P. Frahm, "Some Novel Delta-Function Identities," *Am. J. Phys.*, **51**, 826 (1983).

[112] J. M. Aguirregabiria, A. Hernández, and M. Rivas, "Delta-Function Converging Sequences," *Am. J. Phys.*, **70**, 180 (2002).

Material Properties, Relaxation, and Screening

[113] R. M. Pope and E. S. Fry, "Absorption Spectrum (380–700 nm) of Pure Water. II. Integrating Cavity Measurements," *Appl. Opt.*, **36**, 8710 (97).

[114] N. E. Bengtsson and T. Ohlsson, "Microwave Heating in the Food Industry," *Proc. IEEE*, **62**, 44 (1974).

[115] W. M. Saslow and G. Wilkinson, "Expulsion of Free Electronic Charge from the Interior of a Metal," *Am. J. Phys.*, **39**, 1244 (1971).

[116] N. Ashby, "Relaxation of Charge Imbalances in Conductors," *Am. J. Phys.*, **43**, 553 (1975).

[117] H. C. Ohanian, "On the Approach to Electro- and Magneto-Static Equilibrium," *Am. J. Phys.*, **51**, 1020 (1983).

[118] E. J. Bochove and J. F. Walkup, "A Communication on Electrical Charge Relaxation in Metals," *Am. J. Phys.*, **58**, 131 (1990).

[119] S. Fahy, C. Kittel, and S. G. Louie, "Electromagnetic Screening by Metals," *Am. J. Phys.*, **56**, 989 (1988).

[120] A. R. von Hippel, *Dielectrics and Waves*, MIT Press, Cambridge, MA, 1954.

[121] S. A. Korff and G. Breit, "Optical Dispersion," *Rev. Mod. Phys.*, **4**, 471 (1932).

[122] G. Breit, "Quantum Theory of Dispersion," *Rev. Mod. Phys.*, **4**, 504 (1932), and *ibid.*, **5**, 91 (1933).

[123] R. Ladenburg, "Dispersion in Electrically Excited Gases,", *Rev. Mod. Phys.*, **5**, 243 (1933).

[124] I. H. Malitson, "Interspecimen Comparison of the Refractive Index of Fused Silica," *J. Opt. Soc. Am.*, **55**, 1205 (1965).

[125] P. B. Johnson and R. W. Christy, "Optical Constants of the Noble Metals," *Phys. Rev.*, **B-6**, 4370 (1972).

[126] G. M. Hale and M. R. Querry, "Optical Constants of Water in the 200-nm to 200-μm Wavelength Region," *Appl. Opt.*, **12**, 555 (1973).

[127] W. J. Tropf, M. E. Thomas, and T. J. Harris, "Properties of Crystal and Glasses," in M. Bass, Ed., *Handbook of Optics*, vol. II, McGraw-Hill, New York, 1995. And, R. A. Paquin, "Properties of Metals," *ibid.*.

[128] A. K. Jonscher, "Dielectric Relaxation in Solids," *J. Phys. D: Appl. Phys.*, **32** R57 (1999).

[129] J. J. Makosz and P. Urbanowicz, "Relaxation and Resonance Absorption in Dielectrics," *Z. Naturforsch.*, **57-A**, 119 (2002).

Group Velocity, Energy Velocity, Momentum, and Radiation Pressure

[130] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of Continuous Media*, 2/e, Elsevier Science, Burlington, MA, 1985.

[131] W. Pauli, *Optics and the Theory of Electrons*, Dover Publications, Mineola, NY, 1973.

[132] H. Pelzer, "Energy Density of Monochromatic Radiation in a Dispersive Medium," *Proc. Roy. Soc. London. Ser. A*, **208**, 365 (1951).

[133] M. A. Biot, "General Theorems on the Equivalence of Group Velocity and Energy Transport," *Phys. Rev.*, **105**, 1129 (1957).

[134] R. Loudon, "The Propagation of Electromagnetic Energy through an Absorbing Dielectric," *J. Phys. A*, **3**, 233 (1970).

[135] J. P. Gordon, "Radiation Forces and Momenta in Dielectric Media," *Phys. Rev.*, **A-8**, 14 (1973).

[136] M. V. Berry, "On the Ubiquity of the Sine Wave," *Am. J. Phys.*, **43**, 91 (1975).

[137] K. E. Oughstun and S. Shen, "Velocity of Energy Transport for a Time-Harmonic Field in a Multiple-Resonance Lorentz Medium," *J. Opt. Soc. Am.*, **B-5**, 2395 (1988).

[138] G. C. Sherman and K. E. Oughstun, "Energy Velocity Description of Pulse Propagation in Absorbing, Dispersive Dielectrics," *J. Opt. Soc. Am.*, **B-12**, 229 (1995).

[139] G. Diener, "Energy Transport in Dispersive Media and Superluminal Group Velocities," *Phys. Lett.*, **A-235**, 118 (1997).

[140] K. M. Awati and T. Howes, "Question #52. Group Velocity and Energy Propagation," *Am. J. Phys.*, **64**, 1353 (1996), and the answers by K. T. McDonald, *ibid.*, **66**, 656 (1998); C. S. Helrich, *ibid.*, p. 658; R. J. Mather, *ibid.*, p. 659; S. Wong and D. Styer, *ibid.*, p. 659; and A. Bers, *ibid.*, **68**, 482 (2000).

[141] J. Peatross, S. A. Glasgow, and M. Ware, "Average Energy Flow of Optical Pulses in Dispersive Media," *Phys. Rev. Lett.*, **84**, 2370(2000).

[142] S. Glasgow, M. Ware, and J. Peatross, "Poynting's Theorem and Luminal Total Energy Transport in Passive Dielectric Media," *Phys. Rev.*, **E-64**, 046610 (2001).

[143] C-G. Huang and Y-Z. Zhang, "Poynting Vector, Energy Density, and Energy Velocity in an Anomalous Dispersion Medium," *Phys. Rev.*, **A-65**, 015802 (2001).

[144] R. Ruppin, "Electromagnetic Energy Density in a Dispersive and Absorptive Material," *Phys. Lett.*, **A-299**, 309 (2002).

[145] D. F. Nelson, "Momentum, Pseudomomentum, and Wave Momentum: Toward Resolving the Minkowski-Abraham Controversy," *Phys. Rev.*, **A-44**, 3985 (1991).

[146] R. H. Romer, "Question #26. Electromagnetic Field Momentum," *Am. J. Phys.*, **63**, 777 (1995), and the answers by K. T. McDonald, *ibid.*, **64**, 15 (1996); F. Rohrlich, *ibid.*, p. 16; B. R. Holstein, *ibid.*, p. 17;

[147] R. Loudon, L. Allen and D. F. Nelson, "Propagation of Electromagnetic Energy and Momentum Through an Absorbing Dielectric," *Phys. Rev.*, **E-55**, 1071 (1997).

[148] R. Loudon, "Theory of the Radiation Pressure on Dielectric Surfaces," J. Mod. Opt., **49**, 821 (2002).

[149] M. Mansuripur, "Radiation Pressure and the Linear Momentum of the Electromagnetic Field," *Opt. Express*, **12**, 5375 (2004).

[150] B. A. Kemp, T. M. Gregorczyk, and J. A. Kong, "Ab Initio Study of the Radiation Pressure on Dielectric and Magnetic Media," *Opt. Express*, **13**, 9280 (2005).

[151] S. Stallinga, "Energy and Momentum of Light in Dielectric Media," *Phys. Rev.*, **E-73**, 026606 (2006).

[152] M. Scalora, et al., "Radiation Pressure of Light Pulses and Conservation of Linear Momentum in Dispersive Media," *Phys. Rev.*, **E-73**, 056604 (2006).

[153] R. N. C. Pfeifer, et al., "Momentum of an electromagnetic wave in dielectric media," *Rev. Mod. Phys.*, **79**, 1179 (2007).

Pulse Propagation, Spreading, and Dispersion Compensation

[154] L. Brillouin, *Wave Propagation and Group Velocity*, Academic Press, New York, 1960.

[155] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill, New York, 1962.

[156] G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications*, Van Nostrand, New York, 1948.

[157] A. Ghatak and K. Thyagarajan, *Introduction to Fiber Optics*, Cambridge Univ. Press, Cambridge, 1998.

[158] R. W. Boyd, *Nonlinear Optics* Academic, New York, 2003.

[159] H. M. Nussenzveig, *Causality and Dispersion Relations*, Academic Press, New York, 1972.

[160] V. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas*, 2nd ed., Pergamon Press, New York, 1970.

[161] L. A. Vainshtein, "Propagation of Pulses," *Usp. Fiz. Nauk*, **118**, 339 (1976).

[162] J. Weber, "Phase, Group, and Signal Velocity," *Am. J. Phys.*, **22**, 618 (1954).

[163] M. P. Forrer, "Analysis of Millimicrosecond RF Pulse Transmission," *Proc. IRE*, **46**, 1830 (1958).

[164] R. E. Haskell and C. T. Case, "Transient Signal Propagation in Lossless, Isotropic Plasmas," *IEEE Trans. Antennas Propagat.*, **AP-15**, 458 (1967).

[165] S. Aksornkitti, H. C. S. Hsuan, and K. E. Lonngren, "Dispersion of an Electromagnetic Pulse," *Am. J. Phys.*, **37**, 783 (1969).

[166] R. L. Smith, "The Velocities of Light,", *Am. J. Phys.*, **38**, 978 (1970).

[167] J. Jones, "On the Propagation of a Pulse through a Dispersive Medium," *Am. J. Phys.*, **42**, 43 (1974).

[168] D. York, "Graphical Approach to Dispersion," *Am. J. Phys.*, **43**, 725 (1975).

[169] W. V. Prestwich, "Precise Definition of Group Velocity," *Am. J. Phys.*, **43**, 832 (1975). See also the comment by E. E. Bergmann, *ibid.*, **44**, 890 (1976).

[170] H. M. Bradford, "Propagation and Spreading of a Pulse or Wave Packet," *Am. J. Phys.*, **44**, 1058 (1976).

[171] S. C. Bloch, "Eighth Velocity of Light," *Am. J. Phys.*, **45**, 538 (1977).

[172] H. M. Bradford, "Propagation of a Step in the Amplitude or Envelope of a Pulse or Wave Packet," *Am. J. Phys.*, **47**, 688 (1979).

[173] C. Almeida and A. Jabs, "Spreading of a Relativistic Wave Packet,", *Am. J. Phys.*, **52**, 921 (1984).

[174] J. M. Saca, "Addition of Group Velocities," *Am. J. Phys.*, 173 (1985).

[175] J. A. Lock, "The Temporary Capture of Light by a Dielectric Film," *Am. J. Phys.*, **53**, 968 (1985).

[176] P. C. Peters, "Does a Group Velocity Larger than *c* Violate Relativity?", *Am. J. Phys.*, **56**, 129 (1988).

[177] R. A. Bachman, "Relativistic Phase Velocity Transformation," *Am. J. Phys.*, **57**, 628 (1989).

[178] W-C. Wang and C-M. Chen, "About the Statistical Meaning of Group Velocity," *Am. J. Phys.*, **58**, 744 (1990).

[179] A. Albareda, et al., "Simulation of Dispersion Propagation in Dispersive Media using FFT," *Am. J. Phys.*, **58**, 844 (1990).

[180] F. S. Johnson, "Physical Cause of Group Velocity in Normally Dispersive, Nondissipative Media," *Am. J. Phys.*, **58**, 1044 (1990).

[181] M. Tanaka, "Description of a Wave Packet Propagating in Anomalous Dispersion Media—A New Expression of Propagation Velocity," *Plasma Phys. & Contr. Fusion*, **31** 1049 (1989).

[182] C. J. Gibbins, "Propagation of Very Short Pulses Through the Absorptive and Dispersive Atmosphere," *IEE Proc., Pt. H*, **137**, 304 (1990).

[183] P. Stenius and B. York, "On the Propagation of Transients in Waveguides," *IEEE Ant. Propag. Mag.*, **37**, p. 339, April 1995.

[184] S. L. Dvorak and D. G. Dudley, "Propagation of Ultra-Wide-Band Electromagnetic Pulses Through Dispersive Media," *IEEE Trans. Electrom. Compat.*, **37**, 192 (1995). See also the comment by S. He, *ibid.*, **38**, 202 (1996).

[185] P. Hillion, "Propagation of Electromagnetic Pulses in an Infinite Drude Medium," *IEEE Trans. Electromagn. Compat.*, **39** 361 (1997).

[186] J. Leander, "On the Relation Between the Wavefront Speed and the Group Velocity Concept," *J. Acoust. Soc. Am.*, **100**, 3503 (1996).

[187] P. Loughlin and L. Cohen, "Local Properties of Dispersive Pulses,", *J. Mod. Opt.*, **49**, 2645 (2002).

[188] L. Yang and I. R. Epstein, "Chemical Wave Packet Propagation, Reflection, and Spreading," *J. Phys. Chem.*, **A-106**, 11676 (2002).

[189] S. A. Ramakrishna and A. D. Armour, "Propagating and Evanescent Waves in Absorbing Media," *Am. J. Phys.*, **71**, 562 (2003).

[190] Y. Pinhasi, et al., "Study of Ultrawide-Band Transmission in the Extremely High Frequency (EHF) Band," *IEEE Trans. Antennas Propagat.*, **52**, 2833 (2004).

[191] F. S. Crawford, "Chirped Handclaps," *Am. J. Phys.*, **38**, 378 (1970).

[192] F. S. Crawford, "Douglas Fir Echo Chamber," *Am. J. Phys.*, **38**, 1477 (1970).

[193] F. S. Crawford, "Culvert Whistlers," *Am. J. Phys.*, **39**, 610 (1971).

[194] P. M. Rinard, "Rayleigh, Echoes, Chirps, and Culverts," *Am. J. Phys.*, **40**, 923 (1972).

[195] D. Marcuse, "Pulse Distortion in Single-Mode Fibers," *Appl. Opt.*, **19**, 1653 (1980).

[196] D. Marcuse, "Pulse Distortion in Single-Mode Fibers. Part 2," *Appl. Opt.*, **20**, 2969 (1981).

[197] D. Marcuse, "Pulse Distortion in Single-Mode Fibers. 3: Chirped Pulses," *Appl. Opt.*, **20**, 3573 (1981).

[198] T. L. Koch and R. C. Alferness, "Dispersion Compensation by Active Predistorted Signal Synthesis," *J. Lightwave Tech.*, **LT-3**, 800 (1985).

[199] D. Anderson and M. Lisak, "Analytic Study of Pulse Broadening in Dispersive Optical Fibers," *Phys. Rev.*, **A-35**, 184 (1987).

[200] J. E. Roman and K. A. Winick, "Waveguide Grating Filters for Dispersion Compensation and Pulse Compression," *IEEE J. Quant. Electr.*, **29**, 975 (1993).

[201] B. Jopson and A. Gnauck, "Dispersion Compensation for Optical Fiber Systems," *IEEE Comm. Mag.*, **33**, no.6, 96 (1995).

[202] H. A. Haus, "Optical Fiber Solitons, Their Properties and Uses," *Proc. IEEE*, **81**, 970 (1993).

[203] J. M. Arnold, "Solitons in Communications," *Electr. & Commun. Eng. J.*, **8**, no. 2, 88 (1996).

[204] H. A. Haus, "Group Velocity, Energy, and Polarization Mode Dispersion," *J. Opt. Soc. Am.*, **B-16**, 1863 (1999).

[205] P. Lazaridis, G. Debarge, and P. Gallion, "Discrete Orthogonal Gauss-Hermite Transform for Optical Pulse Propagation Analysis," *J. Opt. Soc. Am.*, **B-20**, 1508 (2003).

[206] ITU-T Recommendation G.652, 6/2005, available from `www.itu.int`.

Precursors

[207] K. E. Oughstun and G. C. Sherman, *Electromagnetic Pulse Propagation in Causal Dielectrics*, Springer-Verlag, Berlin, 1994.

[208] H. G. Baerwald, "Über die Fortpflanzung von Signalen in Dispergierenden Systemen," *Ann. Physik*, **7**, 731 (1930).

[209] P. Pleshko and I. Palócz, "Experimental Observation of Sommerfeld and Brillouin Precursors in the Microwave Domain,", *Phys. Rev. Lett.*, **22**, 1201 (1969).

[210] D. B. Trizna and T. A. Weber, "Brillouin Revisited: Signal Velocity Definition for Pulse Propagation in a Medium with Resonant Anomalous Dispersion," *Radio Sci.*, **17**, 1169 (1982).

[211] R. Albanese, J. Penn, and R. Medina, "Short-Rise-Time Microwave Pulse Propagation Through Dispersive Biological Media," *J. Opt. Soc. Am.*, **A-6**, 1441 (1989).

[212] K. E. Oughstun, “Pulse Propagation in a Linear, Causally Dispersive Medium,” *Proc. IEEE*, **79**, 1379 (1991).

[213] C. M. Balictsis and K. E. Oughstun, “Uniform Asymptotic Description of Ultrashort Gaussian-Pulse Propagation in a Causal, Dispersive Dielectric,”, *Phys. Rev.*, **E-47**, 3645 (1993).

[214] C. M. Balictsis and K. E.Oughstun, “Generalized Asymptotic Description of the Propagated Field Dynamics in Gaussian Pulse Propagation in a Linear, Causally Dispersive Medium,” *Phys. Rev.*, **E-55**, 1910 (1997).

[215] F. J. Ribeiro and M. L. Cohen, “Amplifying Sommerfeld Precursors and Producing a Discontinuous Index of Refraction with Gains and Losses,” *Phys. Rev.*, **E-64**, 046602 (2001).

[216] K. E. Oughstun and N. A. Cartwright, ”Dispersive Pulse Dynamics and Associated Pulse Velocity Measures,” *J. Optics A: Pure Appl. Opt.*, **4**, S125 (2002).

[217] S-H. Choi and U. Österberg, “Observation of Optical Precursors in Water,” *Phys. Rev. Lett.*, **92**, 193903 (2004).

[218] H. Jeong, A. M.C. Dawes, and D. J. Gauthier “Direct Observation of Optical Precursors in a Region of Anomalous Dispersion,” *Phys. Rev. Lett.*, **96**, 143901 (2006).

[219] R. Safian, M. Mojahedi, and C. D. Sarris, “Asymptotic Description of Wave Propagation in an Active Lorentzian Medium,” *Phys. Rev.*, **E-75**, 066611 (2007).

Slow, Fast, and Negative Group Velocity

[220] P. W. Milonni, *Fast Light, Slow Light and Left-Handed Light*, IOP Publishing, London, 2005.

[221] A. Icsevgi and W. E. Lamb, “Propagation of Light Pulses in a Laser Amplifier,” *Phys. Rev.*, **185**, 517 (1969).

[222] C. G. B. Garrett and D. E. McCumber, “Propagation of a Gaussian Light Pulse through an Anomalous Dispersion Medium,” *Phys. Rev.*, **A-1**, 305 (1970).

[223] M. D. Crisp, “Concept of Group Velocity in Resonant Pulse Propagation ,” *Phys. Rev.*, **A-4**, 2104 (1971).

[224] M. D. Crisp, “Propagation of Small-Area Pulses of Coherent Light through a Resonant Medium,” *Phys. Rev.*, **A-1**, 1604 (1971).

[225] L. Casperson and A. Yariv, “Pulse Propagation in a High-Gain Medium,” *Phys. Rev. Lett.*, **26**, 293 (1971).

[226] J. Jones, “On the Propagation of a Pulse Through a Dispersive Medium,” *Am. J. Phys.*, **42**, 43 (1974).

[227] D. Anderson, J. Askne, and M. Lisak, “Wave Packets in an Absorptive and Strongly Dispersive Medium,” *Phys. Rev.*, **A-12**, 1546 (1975).

[228] S. Chu and S. Wong, “Linear Pulse Propagation in an Absorbing Medium ,” *Phys. Rev. Lett.*, **48**, 738 (1982).

[229] B. Ségard and B. Macke, “Observation of Negative Velocity Pulse Propagation,” *Phys. Lett.* **A-109**, 213 (1985).

[230] M. Tanaka, M. Fujiwara, and H. Ikegami, “Propagation of a Gaussian Wave Packet in an Absorbing Medium ,” *Phys. Rev.*, **A-34**, 4851 (1986).

[231] R. Y. Chiao, “Superluminal (but causal) Propagation of Wave Packets in Transparent Media with Inverted Atomic Populations,” *Phys. Rev.*, **A-48**, R34 (1993).

[232] E. L. Bolda, R. Y. Chiao, and J. C. Garrison, “Two Theorems for the Group Velocity in Dispersive Media ,” *Phys. Rev.*, **A-48**, 3890 (1993).

[233] J. M. Deutch and F. E. Low, “Barrier Penetration and Superluminal Velocity,” *Ann. Phys.*, **228**, 184 (1993).

[234] E. L. Bolda, J. C. Garrison, and R. Y. Chiao, “Optical Pulse Propagation at Negative Group Velocities due to a Nearby Gain Line,” *Phys. Rev.*, **A-49**, 2938 (1994).

[235] A. M. Steinberg and R. Y. Chiao, “Dispersionless, Highly Superluminal Propagation in a Medium with a Gain Doublet,”*Phys. Rev.*, **A-49**, 2071 (1994).

[236] Wang Yun-ping and Zhang Dian-lin, “Reshaping, Path Uncertainty, and Superluminal Traveling,” *Phys. Rev.*, **A-52**, 2597 (1995).

[237] R. Y. Chiao, “Population Inversion and Superluminality,” in *Amazing Light*, R. Y. Chiao, ed., Springer, New York, 1996.

[238] Y. Japha and G. Kurizki, “Superluminal Delays of Coherent Pulses in Nondissipative Media: A Universal Mechanism,” *Phys. Rev.*, **A-53**, 586 (1996).

[239] G, Diener, “Superluminal Group Velocities and Information Transfer,” *Phys. Lett.*, **A-223**, 327 (1996).

[240] R. Y. Chiao and A. M Steinberg, “Tunneling Times and Superluminality,” in E. Wolf, ed., *Progr. Optics*, vol. XXXVII, Elsevier, New York, 1997.

[241] Y. Aharonov, B. Reznik, and A. Stern, “Quantum Limitations on Superluminal Propagation,” *Phys. Rev. Lett.*, **81**, 2190 (1998).

[242] M. Büttiker1, H. Thomas, “Front Propagation in Evanescent Media,” *Ann. Phys. (Leipzig)*, **7**, 602 (1998).

[243] G. Nimtz, “Superluminal Signal Velocity,” *Ann. Phys. (Leipzig)*, **7**, 618 (1998).

[244] F. E. Low, “Comments on Apparent Superluminal Propagation,” *Ann. Phys. (Leipzig)*, **7**, 660 (1998).

[245] B. Segev, P. W. Milonni, J. F. Babb, and R. Y. Chiao, “Quantum Noise and Superluminal Propagation,” *Phys. Rev.*, **A-62**, 62, 022114 (2000).

[246] M. A. I. Talukder, Y. Amagishi, and M. Tomita, “Superluminal to Subluminal Transition in the Pulse Propagation in a Resonantly Absorbing Medium,” *Phys. Rev. Lett.*, **86**, 3546 (2000).

[247] L. J. Wang, A. Kuzmich, and A. Dogariu, “Gain-Assisted Superluminal Light Propagation,” *Nature*, **406**, 277 (2000).

[248] A. Dogariu, A. Kuzmich, and L. J. Wang, “Transparent Anomalous Dispersion and Superluminal Light-Pulse Propagation at a Negative Group Velocity,” *Phys. Rev.*, **A-63**, 053806 (2001).

[249] J. Marangos, “Faster Than a Speeding Photon,” *Nature*, **406**, 243 (2000).

[250] K. T. McDonald, “Negative group velocity,” *Am. J. Phys.*, **69**, 607 (2001).

[251] P. Sprangle, J. R. Peñano, and B. Hafizi “Apparent Superluminal Propagation of a Laser Pulse in a Dispersive Medium,” *Phys. Rev.*, **E-64**, 026504 (2001).

[252] T. Sauter, “Superluminal Signals: an Engineer’s Perspective,” *Phys. Lett.*, **A-282**, 145 (2001).

[253] A. D. Jackson A. Lande, and B. Lautrup, “Apparent Superluminal Behavior in Wave Propagation,” *Phys. Rev.*, **A-64**, 044101 (2001).

[254] T. Sauter, “Gaussian Pulses and Superluminality,” *J. Phys. A: Math. Gen.*, **35**, 6743 (2002).

[255] C-G Huang and Y-Z Zhang, “Negative Group Velocity and Distortion of a Pulse in an Anomalous Dispersion Medium,” *J. Opt. A: Pure Appl. Opt.*, **4**, 263 (2002).

[256] R. Y. Chiao and P. W. Milonni, “Fast Light, Slow Light,” *Opt. Photon. News*, p. 27, June 2002.

[257] J. E. Heebner and R. W. Boyd, “ ‘Slow’ and ‘Fast’ Light in Resonator-Coupled Waveguides,” *J. Mod. Opt.*, **49**, 2629 (2002).

[258] G. Nimtz and A. Haibel, “Basics of Superluminal Signals,” *Ann. Phys. (Leipzig)*, **11**, 163 (2002).

[259] W. M. Robertson, J. Ash, and J. M. McGaugh, “Breaking the Sound Barrier: Tunneling of Acoustic Waves Through the Forbidden Transmission Region of a One-Dimensional Acoustic Band Gap Array,” *Am. J. Phys.*, **70**, 689 (2002).

[260] H. Tanaka, et al., “Propagation of Optical Pulses in a Resonantly Absorbing Medium: Observation of Negative Velocity in Rb Vapor,” *Phys. Rev.*, **A-68**, 053801 (2003).

[261] H. Cao, A. Dogariu, and L. J. Wang, “Negative Group Delay and Pulse Compression in Superluminal Pulse Propagation,” *IEEE J. Sel. Top. Quant. Electr.*, **9**, 52 (2003).

[262] B. Macke and B. Ségard, “Propagation of Light-Pulses at a Negative Group Velocity,” *Eur. Phys. J.*, **D 23**, 125 (2003).

[263] M. D. Stenner, D. J. Gauthier, and M. A. Neifeld, “The Speed of Information in a ‘fast-light’ Optical Medium,” *Nature*, **425**, 695 (2003).

[264] H. G. Winful, “Nature of ‘Superluminal’ Barrier Tunneling,” *Phys. Rev. Lett.*, **90**, 023901 (2003).

[265] M. Büttiker and S. Washburn, “Ado About Nothing Much?,” *Nature*, **422**, 271 (2003).

[266] H. G. Winful, “Mechanism for ‘Superluminal’ Tunneling,” *Nature*, **424**, 638 (2003).

[267] Z. M. Zhang and K. Park, “On the Group Front and Group Velocity in a Dispersive Medium Upon Refraction From a Nondispersive Medium,” *J. Heat Transf.*, **126**, 244 (2004).

[268] J. N. Munday and R. H. Henderson, “Superluminal Time Advance of a Complex Audio Signal,” *Appl. Phys. Lett.*, **85**, 503 (2004).

[269] A. Ranfagni, D. Mugnai, and M. A. Vitali, “Real-Time Analysis of the Telegrapher’s Equation for Tunneling Processes,” *Phys. Rev.*, **E-69**, 057603 (2004).

[270] M. D. Stenner, D. J. Gauthier, and M. A. Neifeld, “Fast Causal Information Transmission in a Medium with a Slow Group Velocity,” *Phys. Rev. Lett.*, **94**, 053902 (2005).

[271] Y. Okawachi, et al., “Tunable All-Optical Delays via Brillouin Slow Light in an Optical Fiber,” *Phys. Rev. Lett.*, **94**, 153902 (2005).

[272] J. R. Klauder, “Signal Transmission in Passive Media,” *IEE Proc.–Radar Sonar Navig.*, **152**, 23 (2005).

[273] G. A. Mourou, T. Tajima, and S. V. Bulanov, “Optics in the Relativistic Regime,” *Rev. Mod. Phys.*, **78**, 309 (2006).

[274] W. Guo, “Understanding Subluminal and Superluminal Propagation Through Superposition of Frequency Components,” *Phys. Rev.*, **E-73**, 016605 (2006).

[275] G. M. Gehring, et al., “Observation of Backward Pulse Propagation Through a Medium with a Negative Group Velocity,” *Science*, **312**, 895 (2007).

[276] M. W. Mitchell and R. Y. Chiao, “Causality and Negative Group Delays in a Simple Bandpass Amplifier,” *Am. J. Phys.*, **66**,14 (1998). See also, *Phys. Lett.*, **A-230**, 133 (1997).

[277] J. C. Garrison, M. W. Mitchell, R. Y. Chiao, E. L. Bolda, “Superluminal Signals: Causal Loop Paradoxes Revisited,” *Phys. Lett.*, **A-245**, 19 (1998).

[278] D. Solli, R. Y. Chiao, and J. M. Hickmann, “Superluminal Effects and Negative Group Delays in Electronics, and their Applications,” *Phys. Rev.*, **E-66**, 056601 (2002).

[279] M. Kitano, T. Nakanishi, and K. Sugiyama, “Negative Group Delay and Superluminal Propagation: An Electronic Circuit Approach,” *IEEE J. Sel. Top. Quant. Electr.*, **9**, 43 (2003). 046604 (2001).

[280] T. Nakanishi, K. Sugiyama, and M. Kitano, “Simulation of Slow Light with Electronic Circuits,” *Am. J. Phys.*, **73**, 323 (2005).

[281] S. E. Harris, “Electromagnetically Induced Transparency,” *Phys. Today*, **50**, no.7, 36 (1997).

[282] O. A. Kocharovskaya and Y. I. Khanin, “Coherent Amplification of an Ultrashort Pulse in a Three-Level Medium Without a Population Inversion,” *JETP Lett.*, **48**, 630 (1988).

[283] S. E. Harris, “Lasers Without Inversion—Interference of Lifetime-Broadened Resonances,” *Phys. Rev. Lett.*, **62**, 10336 (1989).

[284] O. Kocharovskaya, “Amplification and Lasing Without Inversion,” *Phys. Rep.*, **219**, 175 (1992).

[285] S. E. Harris, J. E. Field, and A. Kasapi, “Dispersive Properties of Electromagnetically Induced Transparency,” *Phys. Rev.*, **A-46**, R29 (1992).

[286] A. Kasapi, M. Jain, G. Y. Yin, and S. E. Harris, “Electromagnetically Induced Transparency: Propagation Dynamics,” *Phys. Rev. Lett.*, **74**, 2447 (1995).

[287] S. E. Harris, “Electromagnetically Induced Transparency in an Ideal Plasma,” *Phys. Rev. Lett.*, **77**, 5357 (1996).

[288] M. M. Kash, et al., “Ultraslow Group Velocity and Enhanced Nonlinear Optical Effects in a Coherently Driven Hot Atomic Gas,” *Phys. Rev. Lett.*, **82**, 5229 (1999).

[289] L. V. Hau, S. E. Harris, Z. Dutton, and C. H. Behroozi, “Light Speed Reduction to 17 Metres per Second in an Ultracold Atomic Gas,” *Nature*, **397**, 594 (1999).

[290] K. T. McDonald, “Slow Light,” *Am. J. Phys.*, **68**, 293 (2000).

[291] D. F. Phillips, et al., “Storage of Light in Atomic Vapor,” *Phys. Rev. Lett.*, **86**, 783 (2001).

[292] O. Kocharovskaya, Y. Rostovtsev, and M. O. Scully, “Stopping Light via Hot Atoms,” *Phys. Rev. Lett.*, **86**, 628 (2001).

[293] M. D. Lukin and A. Imamoglu, “Controlling Photons Using Electromagnetically Induced Transparency,” *Nature*, **413**, 273 (2001).

[294] T. Pang, “Electromagnetically Induced Transparency,” *Am. J. Phys.*, **69**, 604 (2001).

[295] A. V. Turukhin, et al., “Observation of Ultraslow and Stored Light Pulses in a Solid,” *Phys. Rev. Lett.*, **88**, 23602 (2002).

[296] R. W. Boyd and D. J. Gauthier, “Slow and Fast Light,” in *Progress in Optics*, vol.43, E. Wolf, ed., Elsevier, Amsterdam, 2002.

[297] C. L. G. Alzar, M. A. G. Marinez, and P. Nussenzveig, “Classical Analog of Electromagnetically Induced Transparency,” *Am. J. Phys.*, **70**, 37 (2002).

[298] A. V. Turukhin et al., “Observation of Ultraslow and Stored Light Pulses in a Solid,” *Phys. Rev. Lett.*, **88**, 023602 (2002).

[299] A. G. Litvak and M. D. Tokman, “Electromagnetically Induced Transparency in Ensembles of Classical Oscillators,” *Phys. Rev. Lett.*, **88**, 095003 (2002).

[300] M. Bajcsy, A. S. Zibrov, and M. D. Lukin, “Stationary Pulses of Light in an Atomic Medium,” *Nature*, **426**, 638 (2003).

[301] M. O. Scully, “Light at a Standstill,” *Nature*, **426**, 610 (2003).

[302] M. S. Bigelow, N. N. Lepeshkin, R. W. Boyd “Observation of Ultraslow Light Propagation in a Ruby Crystal at Room Temperature,” *Phys. Rev. Lett.*, **90**, 113903 (2003).

[303] F. L. Kien, J. Q. Liang, and K. Hakuta, “Slow Light Produced by Far-Off-Resonance Raman Scattering,” *IEEE J. Sel. Top. Quant. Electr.*, **9**, 93 (2003).

[304] A. M. Akulshin, et al., “Light Propagation in an Atomic Medium with Steep and Sign-Reversible Dispersion,” *Phys. Rev.*, **A-67**, 011801(R) (2003).

[305] M. D. Lukin, “Trapping and Manipulating Photon States in Atomic Ensembles,” *Rev. Mod. Phys.*, **75**, 457 (2003).

[306] M. F. Yanik and S. Fan, “Stopping Light All Optically,” *Phys. Rev. Lett.*, **92**, 083901 (2004).

[307] D. D. Smith, et al., “Coupled-Resonator-Induced Transparency,” *Phys. Rev.*, **A-69**, 063804 (2004).

[308] M. Fleischhauer, A. Imamoglu, and J. P. Marangos “Electromagnetically Induced Transparency: Optics in Coherent Media,” *Rev. Mod. Phys.*, **77**, 633 (2005).

[309] R. W. Boyd, D. J. Gauthier and A. L. Gaeta “Applications of Slow Light in Telecommunications,” *Opt. Photon. News*, April 2006, p. 19.

[310] H. A. Haus and E. P. Ippen, “Group Velocity of Solitons,” *Opt. Lett.*, **26**, 1654 (2001).

[311] Joe T. Mok, et al., “Dispersionless Slow Light Using Gap Solitons,” *Nature Phys.*, **2**, 775 (2006).

[312] D. J. Gauthier, “Solitons Go Slow,”, *Nature Photonics*, **1**, 92 (2007).

Chirp Radar and Pulse Compression

[313] M. I. Skolnik, ed., *Radar Handbook*, McGraw-Hill, New York, 1970.

[314] M. I. Skolnik, *Introduction to Radar Principles*, McGraw-Hill, New York, 1980.

[315] C. E. Cook and M. Bernfeld, *Radar Signals*, Artech House, Boston, 1993.

[316] N. Levanon and E. Mozeson, *Radar Signals*, Wiley, New York, 2004.

[317] M. A. Richards, *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005.

[318] C. E. Cook, “Pulse Compression—Key to More Efficient Radar Transmission,”, *Proc. IRE*, **48**, 310 (1960).

[319] J. E. Chin and C. E. Cook, “The Mathematics of Pulse Compression,”, *Sperry Eng. Review*, **12**, 11, October 1959.

[320] J. R. Klauder, A. C. Price, S. Darlington, and W. J. Albersheim, “The Theory and Design of Chirp Radars,” *Bell Syst. Tech. J.*, **39**, 745 (1960).

[321] J. R. Klauder, “The Design of Radar Signals Having Both High Range Resolution and High Velocity Resolution,” *Bell Syst. Tech. J.*, **39**, 809 (1960).

[322] A. B. Boehmer, “Binary Pulse Compression Codes,” *IEEE Trans. Inform.*

[323] G. L. Turin, “An Introduction to Matched Filters,” IRE Trans. Inf. Th., **6**, *311 (1960). See also, G. L. Turin, “An Introduction to Digital Matched Filters,”* Proc. IEEE, **64**, *1092 (1976).*

[324] C. E. Cook, “General Matched Filter Analysis of Linear FM Pulse Compression,”, Proc. IRE, **49**, *831 (1961).*

[325] H. O. Ramp and E. R. Wingrove, “Principles of Pulse Compression,” IEEE Trans. Military Electr., **5**, *109 (1961).*

[326] M. Bernfeld, et al., “Matched Filtering, Pulse Compression, and Waveform Design,” Microwave Journal, *vol. 7, 57, October 1964;* ibid., *vol. 7, 81, November 1964;* ibid., *vol. 7, 70, December 1964;* ibid., *vol. 8, 73, January 1965;*

[327] *E. N. Fowle, et al., "A Pulse Compression System Employing a Linear FM Gaussian Signal,"* Proc. IEEE, **51**, *304 (1963).*

[328] *E. N. Fowle, "The Design of FM Pulse Compression Signals,"* IEEE Trans. Inform. Th., **10**, *61 (1964).*

[329] *R. O. Rowlands, "Detection of a Doppler-Invariant FM Signal by Means of a Tapped Delay Line,"* J. Acoust. Soc. Am., **37**, *608 (1965).*

[330] *J. J. Kroszczyński, "Pulse Compression by Means of Linear-Period Modulation,"* Proc. IEEE, **57**, *1260 (1969). Th.*, **IT-13**, 156 (1967).

[331] W. E. Kock, "Pulse Compression with Periodic Gratings and Zone Plate Gratings," *Proc. IEEE*, **58**, 1395 (1970).

[332] R. A. Altes and E. L. Titlebaum, "Bat Signals as Optimally Doppler Tolerant Waveforms," *J Acoust. Soc. Am.*, **48** 1014 (1970).

[333] J. D. Rhodes, "Matched-Filter Theory for Doppler-Invariant Pulse Compression," *IEEE Trans. Circ. Th.*, **CT-19**, 53 (1972).

[334] S. C. Bloch, "Introduction to Chirp Concepts with a Cheap Chirp Radar," *Am. J. Phys.*, **41**, 857 (1973).

[335] H. M. Gerard, et al., "The Design and Applications of Highly Dispersive Acoustic Surface-Wave Filters," *IEEE Trans. Microwave Theory Tech.*, **MTT-21**, 176 (1973).

[336] D. Hazony and Y. Hazony, "Doppler Invariant Pulse Compressors," *J. Franklin Inst.*, **309**, 215 (1980).

[337] B. L. Lewis and F. F. Kretschmer, "A New Class of Polyphase Pulse Compression Codes and Techniques," *IEEE Trans. Aerosp. Electr. Syst.*, **AES-17**, 364 (1981).

[338] B. L. Lewis and F. F. Kretschmer, "Linear Frequency Modulation Derived Polyphase Pulse Compression Codes," *IEEE Trans. Aerosp. Electr. Syst.*, **AES-18**, 637 (1982).

[339] K. M. El-Shennawy, O. A. Alim, and M. A. Ezz-El-Arab, "Sidelobe Suppression in Low and High Time-Bandwidth Products of Linear FM Pulse Compression Filters," *IEEE Trans. Microwave Theory Tech.*, **MTT-35** 807 (1987). See also the comment by M. K. Roy, *ibid.*, **MTT-36**, 1458 (1988).

[340] S. J. Rabinowitz, et al., "Applications of Digital Technology to Radar," *Proc. IEEE*, **73**, 325 (1985).

[341] H. G. Winful, "Pulse Compression in Optical Fiber Filters," *Appl. Phys. Lett.*, **46**, 527 (1985).

[342] C. E. Cook and W. M. Siebert, "The Early History of Pulse Compression Radar," *IEEE Trans. Aerosp. Electr. Syst.*, **24**, 825 (1988).

[343] S. R. Gottesman, P. G. Grieve, and S. W. Golomb, "A Class of Pseudonoise-Like Pulse Compression Codes," *IEEE Trans. Aerosp. Electr. Syst.*, **28**, 355 (1992).

[344] A. W. Lohmann and D. Mendlovic, "Temporal Filtering with Time Lenses," *Appl. Opt.*, **31**, 6212 (1992).

[345] J. W. Arthur, "Modern SAW-Based Pulse Compression Systems for Radar Applications, Part 1," *Electr. & Commun. Eng. J.*, **7**, 236 (1995); and "Part 2," *ibid.*, **8**, 57 (1996).

[346] P. W. Smith, "Power SAWs," *IEEE Potentials,* **14**, no. 5, 18 (1995).

[347] Y. Tomizawa, et al., "Archaeological Survey Using Pulse Compression Subsurface Radar," *Archeol. Prospect.*, **7**, 241 (2000).

[348] T. Misaridis and J. A. Jensen, "Use of Modulated Excitation Signals in Medical Ultrasound. Part I,", *IEEE Trans. Ultrason., Ferroel., Freq. Contr.*, **52**, 177 (2005); and "Part II,", *ibid.*, p. 192; and "Part III,', *ibid.*, p. 208.

[349] J. Yang and T. K. Sarkar, "Doppler-Invariant Property of Hyperbolic Frequency Modulated Waveforms," *Microw. Opt. Tech. Lett.*, **48**, 1174 (2006).

[350] A. W. Doerry, "Generating Nonlinear FM Chirp Waveforms for Radar," *Sandia National Labs.*, Report SAND2006-5856, 2006, available from `http://www.osti.gov/energycitations`.

[351] J. van Howe and C. Xu, "Ultrafast Optical Signal Processing Based Upon Space-Time Dualities," *J. Lightwave Tech.*, **24**, 2649 (2006).

[352] J. A. Boehm, III and .k S. Debroux, "Distortion of a Received Radar Pulse due to High Target Velocity," *J. Appl. Phys.*, **99**, 124907 (2006).

Negative-Index Media

[353] V. G. Veselago, "The Electrodynamics of Substances with Simultaneously Negative Values of ϵ and μ," *Sov. Phys. Uspekhi*, **10**, 509 (1968).

[354] R. N. Bracewell, "Analogues of An Ionized Medium: Applications to the Ionosphere," *Wireless Eng.* (Iliff & Sons Ltd., London, 1954), p. 320-326.

[355] W. Rotman, "Plasma Simulation by Artificial Dielectrics and Parallel-Plate Media," *IEEE Trans. Antennas Propagat.*, **10**, 82 (1962).

[356] J. B. Pendry, et al., "Extremely Low Frequency Plasmons in Metallic Mesostructures," *Phys. Rev. Lett.*, **76**, 4773(1996).

[357] D. R. Smith, et al., "Loop-wire medium for investigating plasmons at microwave frequencies," *Appl. Phys. Lett.*, **75**, 1425 (1999).

[358] J. B. Pendry, "Magnetism from Conductors and Enhanced Nonlinear Phenomena," *IEEE Trans. Microwave Theory Tech.*, **47**, 2075 (1999).

[359] D. R. Smith, et al., "Composite Medium with Simultaneously Negative Permeability and Permittivity," *Phys. Rev. Lett.*, **84**, 4184 (2000).

[360] J. B. Pendry, "Negative Refraction Makes a Perfect Lens," *Phys. Rev. Lett.*, **85**, 3966 (2000).

[361] D. R. Smith and N. Kroll, "Negative Refractive Index in Left-Handed Materials," *Phys. Rev. Lett.*, **85**, 2933 (2000).

[362] R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental Verification of a Negative Index of Refraction," *Science*, **292**, 7779 (2001).

[363] R. W. Ziolkowski, "Superluminal Transmission of Information through an Electromagnetic Metamaterial," *Phys. Rev.*, **E-63**, 046604 (2001).

[364] R. W. Ziolkowski and E. Heyman, "Wave Propagation in Media Having Negative Permittivity and Permeability," *Phys. Rev.*, **E-64**, 056625 (2001).

[365] I. V. Lindell, et al., "BW Media—Media with Negative Parameters, Capable of Supporting Backward Waves," *Microw. Opt. Tech. Lett.*, **31**, 129 (2001).

[366] K. R. Rao, "Back to Basics: Reflections on Refraction," *Current Sci.*, **81**, 875 (2001).

[367] G. V. Eleftheriades, A. K. Iyer, and P. C. Kremer, "Planar Negative Refractive Index Media Using Periodically L–C Loaded Transmission Lines," *IEEE Trans. Microwave Theory Tech.*, **50**, 2702 (2002).

[368] R. Ruppin, "Surface polaritons of a left-handed medium," *Phys. Lett.*, **A-277**, 61 (2000).

[369] R. Ruppin, "Surface polaritons of a left-handed material slab," *J. Phys.: Condens. Matter*, **13**, 1811 (2001).

[370] F. D. M. Haldane, "Electromagnetic Surface Modes at Interfaces with Negative Refractive Index Make a "Not-Quite-Perfect" Lens," *arXiv*, cond-mat/020420, (2002).

[371] M. Feise, P. J. Bevelacqua, and J. B. Schneider, "Effects of surface waves on the behavior of perfect lenses," *Phys. Rev.*, **B-66**, 035113 (2002).

[372] P. R. Berman, "Goos-Hänchen shift in negatively refractive media," *Phys. Rev.*, **E-69**, 066167 (2004).

[373] N. Garcia and M. Nieto-Vesperinas, "Left-Handed Materials Do Not Make a Perfect Lens," *Phys. Rev. Lett.*, **88**, 207403 (2002); and *ibid.*, **90**, 229903 (2003); and Pendry's reply, *ibid.*, **91**, 099701 (2003).

[374] M. W. McCall, A. Lakhtakia, and W. S. Weiglhofer, "The Negative Index of Refraction Demystified," *Eur. J. Phys.*, **23**, 353 (2002).

[375] D. R. Smith, et al., "Determination of effective parameters and permeability of metamaterials from reflection and transmission coefficients," *Phys. Rev.*, **B-65**, 195104 (2002).

[376] Z. Ye, "Optical transmission and reflection of perfect lenses by left handed materials," *Phys. Rev.*, **B-67**, 193106 (2003).

[377] D. R. Smith, et al., "Limitations on subdiffraction imaging with a negative refractive index slab," *Appl. Phys. Lett.*, **82**, 1506 (2003).

[378] M. Mojahedi, et al., "Abnormal Wave Propagation in Passive Media," *IEEE J. Sel. Top. Quant. Electr.*, **9**, 30 (2003).

[379] C. R. Simovski and B. Sauviac, "On focusing left-handed materials by arbitrary layers," *Microw. Opt. Tech. Lett.*, **39**, 64 (2003).

[380] X. S. Rao and C. K. Ong, "Amplification of evanescent waves in a lossy left-handed material slab," *Phys. Rev.*, **B-68**, 113103 (2003).

[381] X. S. Rao and C. K. Ong, "Subwavelength imaging by a left-handed material superlens," *Phys. Rev.*, **E-68**, 067601 (2003).

[382] G. Gómez-Santos, "Universal Features of the Time Evolution of Evanescent Modes in a Left-Handed Perfect Lens," *Phys. Rev. Lett.*, **90**, 077401 (2003).

[383] C. G. Parazzoli, et al., "Experimental Verification and Simulation of Negative Index of Refraction Using Snell's Law," *Phys. Rev. Lett.*, **90**, 107401 (2003).

[384] A. A. Houck, J. B. Borck, and I. L. Chuang, "Experimental Observations of a Left-Handed Material That Obeys Snell's Law," *Phys. Rev. Lett.*, **90**, 137401 (2003).

[385] J. B. Pendry and S. A. Ramakrishna, "Focusing light using negative refraction," *J. Phys.: Condens. Matter*, **15**, 6345 (2003).

[386] S. A. Cummer, "Simulated causal subwavelength focusing by a negative refractive index slab," *Appl. Phys. Lett.*, **82**, 1503 (2003).

[387] S. Foteinopoulou and C. M. Soukoulis, "Negative refraction and left-handed behavior in two-dimensional photonic crystals," *Phys. Rev.*, **B-67**, 235107 (2003).

[388] E. Cubukcu, et al., "Subwavelength Resolution in a Two-Dimensional Photonic-Crystal-Based Superlens," *Phys. Rev. Lett.*, **91**, 207401 (2003).

[389] E. Cubukcu, et al., "Electromagnetic waves: Negative refraction by photonic crystals," *Nature*, **423**, 604 (2003).

[390] P. V. Parimi, "Photonic crystals: Imaging by flat lens using negative refraction," *Nature*, **426**, 404 (2003).

[391] C. Luo, et al., "Subwavelength imaging in photonic crystals," *Phys. Rev.*, **B-68**, 045115 (2003).

[392] A. Grbic and G. V. Eleftheriades, "Subwavelength Focusing Using a Negative-Refractive-Index Transmission Line Lens," *IEEE Ant. Wireless Prop. Lett.*, **2**, 186 (2003).

[393] A. Grbic and G. V. Eleftheriades, "Overcoming the Diffraction Limit with a Planar Left-Handed Transmission-Line Lens," *Phys. Rev. Lett.*, **92**, 117403 (2004).

[394] C. Caloz and T. Itoh, "Transmission line approach of left-handed (LH) materials and microstrip implementation of an artificial LH transmission line," *IEEE Trans. Antennas Propagat.*, **52**, 1159 (2004).

[395] J. B. Pendry, "A Chiral Route to Negative Refraction," *Science*, **306**, 1353 (2004).

[396] D. R. Smith, J. B. Pendry, and M. C. K. Wiltshire, "Metamaterials and Negative Refractive Index," *Science*, **305**, 788 (2004).

[397] M. Nieto-Vesperinas, "Problem of image superresolution with a negative-index slab," *J. Opt. Soc. Am.*, **A-21**, 491 (2004).

[398] K. Y. Bliokh and Y. P. Bliokh, "What are the left-handed media and what is interesting about them?", *Phys.-Uspekhi*, **47**, 393 (2004).

[399] J. F. Woodley and M. Mojahedi, "Negative Group Velocity and Group Delay in Left-Handed Media," *Phys. Rev.*, **E-70**, 046603 (2004).

[400] S. Dutta Gupta, R. Arun, and G. S. Agarwal, "Subluminal to Superluminal Propagation in a Left-Handed Medium," *Phys. Rev.*, **B-69**, 113104 (2004).

[401] O. F. Siddiqui, et al., "Time-Domain Measurement of Negative Group Delay in Negative-Refractive-Index Transmission-Line Metamaterials," *IEEE Trans. Microwave Theory Tech.*, **52**, 1449 (2004).

[402] R. A. Depine and A. Lakhtakia, "A New Condition to Identify Isotropic Dielectric-Magnetic Materials Displaying Negative Phase Velocity," *Microw. Opt. Technol. Lett.,* **41**, 315 (2004).

[403] X. Chen, et al., "Robust method to retrieve the constitutive effective parameters of metamaterials," *Phys. Rev.*, **E-70**, 016608 (2004).

[404] H. Cory and C. Zach, "Wave Propagation in Metamaterial Multi-Layered Structures," *Microw. Opt. Tech. Lett.*, **40**, 460 (2004).

[405] V. M. Agranovitch, et al., "Linear and nonlinear wave propagation in negative refraction metamaterials," *Phys. Rev.*, **B-69**, 165112 (2004).

[406] J. B. Pendry and D. R. Smith, "Reversing Light with Negative Refraction," *Phys. Today*, **57**, no. 6, 37 (2004).

[407] X. Chen and C-F. Li, "Lateral shift of the transmitted light beam through a left-handed slab," *Phys. Rev.*, **E-69**, 066617 (2004).

[408] X. Huang and W. L. Schaich, "Wave packet propagation into a negative index medium," *Am. J. Phys.*, **72**, 1232 (2004).

[409] S. A. Ramakrishna, "Physics of negative refractive index materials," *Rep. Prog. Phys.*, **68**, 449 (2005).

[410] J. R. Thomas and A. Ishimaru, "Wave Packet Incident on Negative-Index Media," *IEEE Trans. Antennas Propagat.*, **53**, 1591 (2005).

[411] N. Engheta and R. W. Ziolkowski, "A Positive Future for Double-Negative Metamaterials," *IEEE Trans. Microwave Theory Tech.*, **53**, 1535 (2005).

[412] C. Fu, Z. M. Zhang, and P. N. First, "Brewster angle with a negative-index material," *Appl. Opt.*, **44**, 3716 (2005).

[413] S. A. Ramakrishna and O. J. F. Martin, "Resolving the wave vector in negative refractive index media," *Opt. Lett.*, **30**, 2626 (2005).

[414] G. V. Eleftheriades and K. G. Balmain, *Negative Refraction Metamaterials: Fundamental Principles and Applications*, Wiley-IEEE Press, New York, 2005.

[415] A. D. Boardman, N, King, and L. Velasco, "Negative Refraction in Perspective," *Electromagnetics*, **25**, 365 (2005).

[416] J. J. Chen, et al., "Role of evanescent waves in the positive and negative Goos-Hänchen shifts with left-handed material slabs,", *J. Appl. Phys.*, **98**, 094905 (2005).

[417] T. Koschny, R. Moussa, and C. M. Soukoulis, "Limits on the amplification of evanescent waves of left-handed materials," *arXiv*, cond-mat/0504349, (2005).

[418] K. Aydin, I. Bulu, and E. Ozbay, "Verification of Impedance Matching at the Surface of Left-handed Materials," *Microw. Opt. Tech. Lett.*, **48**, 2548 (2006).

[419] B-I. Popa and S. A. Cummer, "Direct measurement of evanescent wave enhancement inside passive metamaterials," *Phys. Rev.*, **E-73**, 016617 (2006).

[420] T. J. Cui, et al., "Experiments on evanescent-wave amplification and transmission using metamaterial structures," *Phys. Rev.*, **B-73**, 245119 (2006).

[421] V. Veselago, et al., "Negative Refractive Index Materials," *J. Comput. Theor. Nanosci.*, **3**, 1 (2006).

[422] G. Dolling, et al., "Simultaneous Negative Phase and Group Velocity of Light in a Metamaterial," *Science*, **312**, 892 (2006).

[423] Y-F. Chen, P. Fischer, and F. W. Wise, "Sign of the Refractive Index in a Gain Medium with Negative Permittivity and Permeability," *J. Opt. Soc. Am.*, **B-23**, 45 (2006).

[424] C. M. Soukoulis, "Bending Back Light: The Science of Negative Index Materials," *Opt. Photon. News*, **17**, no. 6, 16 (2006).

[425] C. M. Soukoulis, M. Kafesaki, and E. N. Economou, "Negative-Index Materials: New Frontiers in Optics," *Adv. Mater.*, **18**, 1941 (2006).

[426] W. J. Padilla, D. R. Smith, and D. N. Basov, "Spectroscopy of metamaterials from infrared to optical frequencies," *J. Opt. Soc. Am.*, **B-23**, 404 (2006).

[427] C. Caloz and T. Itoh, *Electromagnetic Metamaterials*, Wiley-Interscience, Hoboken, NJ, 2006.

[428] A. V. Kats, et al., "Left-Handed Interfaces for Electromagnetic Surface Waves," *Phys. Rev. Lett.*, **98**, 073901 (2007).

[429] B. A. Kemp, J. A. Kong, and T. M. Grzegorczyk, "Reversal of Wave Momentum in Isotropic Left-Handed Media," *Phys. Rev.*, **A-75**, 053810 (2007).

[430] W. J. Padilla, "Group theoretical description of artificial electromagnetic metamaterials," *Opt. Expr.*, **15**, 1639 (2007).

[431] N. Engheta, "Circuits with Light at Nanoscales: Optical Nanocircuits Inspired by Metamaterials," *Science*, **317**, 1698 (2007).

[432] A. Alù, et al., "Single-negative, Double-Negative, and Low-Index Metamaterials and their Electromagnetic Applications," *IEEE Antennas Propagat. Mag.*, **49**, no. 1, 23 (2007).

[433] A. A. Tretyakov and S. I. Maslovsky, "Veselago Materials: What is Possible and Impossible about the Dispersion of the Constitutive Parameters," *IEEE Antennas Propagat. Mag.*, **49**, no. 1, 37 (2007).

[434] G. V. Eleftheriades, "Enabling RF/Microwave Devices Using Negative-Reflective-Index Transmission-Line (NRI-TL) Metamaterials," *IEEE Antennas Propagat. Mag.*, **49**, no.2, 34 (2007).

Electromagnetism in Moving Media

[435] A. Einstein, "Zur Elektrodynamik Bewegter Körper," *Ann. Physik* (Leipzig), **17**, 891 (1905). Reprinted in [436].

[436] *The Principle of Relativity,* A Collection of Original Memoirs on the Special and General Theory of Relativity, by H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, Traslated by W. Perrett and G. B. Jeffery, Dover Publications, New York, 1952.

[437] J. Van Bladel, *Relativity and Engineering*, Springer-Verlag, Berlin, 1984.

[438] C. T. Tai, "A Study of Electrodynamics of Moving Media," *Proc. IEEE*, **52**, 685 (1964).

[439] C. Yeh, "Reflection and Transmission of Electromagnetic Waves by a Moving Dielectric Medium," *J. Appl. Phys.*, **36**, 3513 (1965).

[440] C. Yeh, "Reflection and Transmission of Electromagnetic Waves by a Moving Plasma Medium," *J. Appl. Phys.*, **37**, 3079 (1966).

[441] C. Yeh and K. F. Casey, "Reflection and Transmission of Electromagnetic Waves by a Moving Dielectric Slab," *Phys. Rev.*, **144**, 665 (1966).

[442] V. P. Pyati, "Reflection and Refraction of Electromagnetic Waves by a Moving Dielectric Medium," *J. Appl. Phys.*, **38**, 652 (1967).

[443] T. Shiozawa, K. Hazama, and N. Kumagai, "Reflection and Transmission of Electromagnetic Waves by a Dielectric Half-Space Moving Perpendicular to the Plane of Incidence," *J. Appl. Phys.*, **38**, 4459 (1967).

[444] C. T. Tai, "Present View on Electrodynamics of Moving Media," *Radio Sci.*, **2**, 245 (1967).

[445] J. F. Holmes and A. Ishimaru, "Relativistic Communication Effects Associated with Moving Space Antennas," *IEEE Trans. Antennas Propagat.*, **AP-17**, 484 (1967).

[446] P. Daly and H. Gruenber, "Energy Relations for Plane Waves Reflected from Moving Media," *J. Appl. Phys.*, **38**, 4486 (1967).

[447] D. Censor, "Scattering of a Plane Wave at a Plane Interface Separating Two Moving Media," *Radio Sci.*, **4**, 1079 (1969).

[448] W. J. Noble and C. K. Ross, "The Reflection and Transmission of Electromagnetic Waves by a Moving Dielectric Slab," *Am. J. Phys.*, **37**, 1249 (1969), and *ibid.*, p.1253.

[449] R. G. Newburgh and T. E. Phipps, Jr., "Brewster Angle and the Einstein Velocity Addition Theorem," *Am. J. Phys.*, **39**, 1079 (1971).

[450] I. Lerche, "Reflection and Refraction of Light from a Moving Block of Glass," *Phys. Rev.*, **D-11**, 740 (1975).

[451] M. Saca, "Brewster Angle in a Semi-Infinite Dielectric Moving Perpendicularly to the Interface," *Am. J. Phys.*, **48**, 237 (1980).

[452] K. Tanaka, "Reflection and Transmission of Electromagnetic Waves by a Linearly Accelerated Dielectric Slab," *Phys. Rev.*, **A-25**, 385 (1982).

[453] J. R. Van Meter, S. Carlip, and F. V. Hartemann, "Reflection of Plane Waves from a Uniformly Accelerating Mirror," *Am. J. Phys.*, **69**, 783 (2001).

[454] J. P. Costella, B. H. J. McKellar, and A. A. Rawlinson, "The Thomas Rotation," *Am. J. Phys.*, **69**, 837 (2001).

[455] H. L. Berk, K. Chaicherdsakul, and T. Udagawa, "The Proper Lorentz Transformation Operator $e^{L} = e^{-\boldsymbol{\omega}\cdot\boldsymbol{S}-\boldsymbol{\xi}\cdot\boldsymbol{K}}$: Where's It Going, What's the Twist?," *Am. J. Phys.*, **69**, 996 (2001).

[456] F. J. Tischer, "Doppler Phenomena in Space Communications," *IRE Trans. Comm. Syst.*, **7**, 25 (1959).

[457] U. Leonhardt and P. Piwnicki, "Relativistic Effects of Light in Moving Media with Extremely Low Group Velocity," *Phys. Rev. Lett.*, **84**, 822 (2000).

Ewald-Oseen Extinction Theorem

[458] P. P. Ewald, "Zur Begründung der Kristalloptik," *Ann. Physik*, Ser. 4, **49**, 1 (1916). See also, *Fortsch. Chem. Phys. Phy. Chem.*, Ser. B, **18**, 491 (1925).

[459] C. W. Oseen, "Über die Wechselwirkung zwischen zwei elektrischen Dipolen and über die Drehung der Polarisationsebene in Kristallen und Flüssigkeiten," *Ann. Physik*, Ser. 4, **48**, 1 (1915).

[460] L. Rosenfeld, *Theory of Electrons*, North Holland Publishing Comp., Amsterdam, 1951.

[461] A. S. Pine, "Self-Consistent Field Theory of Linear and Nonlinear Crystalline Dielectrics Including Local-Field Effects," *Phys. Rev.*, **139**, no. 3A, A901 (1965).

[462] P. P. Ewald, "Crystal Optics for Visible Light and X-Rays," *Reviews of Modern Physics*, **37**, 46 (1965).

[463] V. A. Kizel, "Modern Status of the Theory of Light Reflection," *Sov. Phys. Uspekhi*, **10**, 485 (1968).

[464] J. J. Sein, "A Note on the Ewald-Oseen Extinction Theorem," *Opt. Commun.*, **2**, 170 (1970).

[465] R. K. Bullough, "Many-Body Optics III. The Optical Extinction Theorem," *J. Phys. A*, **3**, 708 (1970).

[466] J. J. Sein, "Boundary Conditions in the Exciton Absorption Region," *J. Opt. Soc. Am.*, **62**, 1037 (1972).

[467] J. J. Sein, "Optics of Polaritons in Bounded Media," *Phys. Rev.*, **B-6**, 2482 (1972).

[468] E. Lalor and E. Wolf, "Exact Solution of the Equations of Molecular Optics for Refraction and Reflection of an Electromagnetic Wave on a Semi-Infinite Dielectric," *J. Opt. Soc. Am.*, **62**, 1165 (1972).

[469] D. N. Pattanyak and E. Wolf, "General Form and a New Interpretation of the Ewald-Oseen Extinction Theorem," *Opt. Commun.*, **6**, 217 (1972).

[470] J. De Goede and P. Mazur, "On the Extinction Theorem in Electrodynamics," *Physica*, **58**, 568 (1972).

[471] J. J. Sein, "General Extinction Theorems," *Opt. Commun.*, **14**, 157 (1975).

[472] J. Van Kranendonk and J. E. Sipe, "Foundations of the Macroscopic Electromagnetic Theory of Dielectric Media," in *Progress in Optics*, vol. XV, E. Wolf, ed., North-Holland Publishing Co., Amsterdam, 1977.

[473] D. Dialetis, "Equivalence of the Ewald-Oseen Extinction Theorem as a Nonlocal Boundary-Value Problem with Maxwell's Equations and Boundary Conditions," *J. Opt. Soc. Am.*, **68**, 602 (1978).

[474] D. E. Aspnes, "Local-Field Effects and Effective-Medium Theory: A Microscopic Perspective," *Am. J. Phys.*, **50**, 704 (1982).

[475] A. T. Friberg and E. Wolf, "Angular Spectrum Representation of Scattered Electromagnetic Fields," *J. Opt. Soc. Am.*, **73**, 26 (1983).

[476] J. J. Sein, "Solutions to Time-Harmonic Maxwell Equations with a Hertz Vector," *Am. J. Phys.*, **57**, 834 (1989).

[477] G. P. M. Poppe and C. M. J. Wijers, "Exact Solution of the Optical Response of Thick Slabs in the Discrete Dipole Approach," *Physica B*, **167**, 221 (1990).

[478] R. K. Bullough and F. Hynne, "Ewald's Optical Extinction Theorem," in *P. P. Ewald and his Dynamical Theory of X-ray Diffraction*, D. W. Cruickshank and H. J. Juretschke, eds., Oxford Univ. Press, New York, 1992.

[479] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, vol.1, Addison-Wesley, Reading, MA, 1963.

[480] F. L. Markley, "The Index of Refraction," *Am. J. Phys.*, **40**, 1799 (1972).

[481] K. S. Kunz and E. Gemoets, "A Simple Model to Explain the Slowing Down of Light in a Crystalline Medium," *Am. J. Phys.*, **44**, 264 (1976).

[482] N. E. Hill, "Reflection and Transmission in Terms of Polarization," *Am. J. Phys.*, **48**, 752 (1980).

[483] R. K. Wangsness, "Effect of Matter on the Phase Velocity of an Electromagnetic Wave," *Am. J. Phys.*, **49**, 950 (1981).

[484] G. C. Reali, "Exact Solution of the Equations of Molecular Optics for Refraction and Reflection of an Electromagnetic Wave on a Semi-Infinite Dielectric," *J. Opt. Soc. Am.*, **72**, 1421 (1982).

[485] G. C. Reali, "Reflection from Dielectric Materials," *Am. J. Phys.*, **50**, 1133 (1982).

[486] M. Schwartz, *Principles of Electrodynamics*, Dover Publications, New York, 1987.

[487] G. C. Reali, "Reflection, Refraction, and Transmission of Plane Electromagnetic Waves from a Lossless Dielectric Slab," *Am. J. Phys.*, **60**, 532 (1992).

[488] M. B. James and D. J. Griffiths, "Why the Speed of Light is Reduced in a Transparent Medium," *Am. J. Phys.*, **60**, 309 (1992). See also Comment by J. B. Diamond, *ibid.*, **63**, 179 (1995).

[489] B. G. de Grooth, "Why Is the Propagation Velocity of a Photon in a Transparent Medium Reduced?" *Am. J. Phys.*, **65**, 1156 (1997).

[490] M. Mansuripur, "The Ewald-Oseen Extinction Theorem," *Optics & Photonics News*, **9** (8), 50 (1998). Reprinted in Ref. [1155].

[491] H. Fearn, D. F. V. James, and P. W. Milonni, "Microscopic Approach to Reflection, Transmission, and the Ewald-Oseen Extinction Theorem," *Am. J. Phys.*, **64**, 986 (1996). See also Comment by H. J. Juretschke, *ibid.*, **67**, 929 (1999).

[492] V. C. Ballenegger and T. A. Weber, “The Ewald-Oseen Extinction Theorem and Extinction Lengths,” *Am. J. Phys.*, **67**, 599 (1999).

[493] H. M. Lai, Y. P. Lau, and W. H. Wong, “Understanding Wave Characteristics via Linear Superposition of Retarded Fields,” *Am. J. Phys.*, **70**, 173 (2002).

Near-Field Optics

[494] S. V. Sukhov and K. V. Krutitsky, “Discrete Structure of Ultrathin Dielectric Films and their Surface Optical Properties,” *Phys. Rev.*, **B-65**, 115407 (2002).

[495] H. F. Arnoldus and J. T. Foley, “Uniform Asymptotic Approximation of the Evanescent Part of the Green’s Tensor,” *Opt. Commun.*, **207**, 7 (2002).

[496] T. Setälä, M. Kaivola, and A. T. Friberg, “Evanescent and Propagating Electromagnetic Fields in Scattering from Point-Dipole Structures,” *J. Opt. Soc. Am.*, **A-18**, 678 (2001). See also, *ibid.*, **A-19**, 1449 (2002), and M. Xiao, *ibid.*, **A-19**, 1447 (2002).

[497] C. Girard, C. Joachim, and S. Gauthier, “The Physics of the Near Field,” *Rep. Progr. Phys.*, **63**, 657 (2000).

[498] A. Lakhtakia and W. S. Weiglhofer, “Evanescent Plane Waves and the Far Field: Resolution of a Controversy,” *J. Mod. Opt.*, **47**, 759 (2000).

[499] M. Ohtsu and H. Hori, *Near-Field Nano-Optics*, Kluwer, New York, 1999.

[500] A. V. Shchegrov and P. S. Carney, “Far-Field Contribution of Evanescent Modes to the Electromagnetic Green Tensor,” *J. Opt. Soc. Am.*, **A-16**, 2583 (1999).

[501] O. Keller, “Attached and Radiated Electromagnetic Fields of an Electric Dipole,” *J. Opt. Soc. Am.*, **B-16**, 835 (1999).

[502] T. Setälä, M. Kaivola, and A. T. Friberg, “Decomposition of the Point-Dipole Field into Homogeneous and Evanescent Parts,” *Phys. Rev.*, **E-59**, 1200 (1999).

[503] E. Wolf and J. T. Foley, “Do Evanescent Waves Contribute to the Far Field?,” *Opt. Lett.*, **23**, 16 (1998).

[504] A. V. Ghiner and G. I. Surdutovich, “Discreteness and Local Fields in Weakly Rarefied Media,” *Phys. Rev.*, **E-56**, 6123 (1997).

[505] K. V. Krutitsky and S. V. Suhov, “Near-Field Effect in Classical Optics of Ultra-Thin Films,” *J. Phys.*, **B-30**, 5341 (1997).

[506] J-J. Greffet and R. Carminati, “Image Formation in Near-Field Optics,” *Progr. Surf. Sci.*, **56**, 133 (1997).

[507] C. Girard and A. Dereux, “Near-Field Optics Theories,” *Rep. Progr. Phys.*, **59**, 657 (1996).

[508] D. Courjon and C. Bainier, “Near Field Microscopy and Near Field Optics,” *Rep. Progr. Phys.*, **57**, 989 (1994).

[509] A. V. Ghiner and G. I. Surdutovich, “Method of Integral Equations and an Extinction Theorem in Bulk and Surface Phenomena in Nonlinear Optics,” *Phys. Rev.*, **A-49**, 1313 (1993).

[510] E. Betzig and J. K. Trautman, “Near-Field Optics: Microscopy, Spectroscopy, and Surface Modification Beyond the Diffraction Limit,” *Science*, **257**, 189 (1992).

[511] L. E. C. van de Leemput and H. van Kempen, “Scanning Tunneling Microscopy,” *Rep. Progr. Phys.*, **55**, 1165 (1992).

[512] E. Betzig, J. K. Trautman, T. D. Harris, J. S. Weiner, and R. L. Kostelak, “Breaking the Diffraction Barrier: Optical Microscopy on a Nanometric Scale,” *Science*, **251**, 1468 (1991).

[513] A. Johner and P. Schaaf, “Calculation of the Reflection Coefficients at Interfaces: A Scattering Approach,” *Phys. Rev.*, **B-42**, 5516 (1990).

[514] U. Dürig, D. W. Pohl, and F. Rohner, “Near-Field Optical-Scanning Microscopy,” *J. Appl. Phys.*, **59**, 3318 (1986).

[515] H. Weyl, “Ausbreitung elektromagnetischer Wellen über einem ebenen Leiter,” *Ann. Physik*, Ser. 4, **60**, 481 (1919).

Total Internal Reflection

[516] F. Goos and H. Hänchen, *Ann. Physik*, (Leipzig), **1**, 333 (1947).

[517] K. Artmann, “Berechnung der Seitenversetzung des totalreflektierten Strahles,” *Ann. Physik*, (Leipzig), **2**, 87 (1948).

[518] J. Fahrenfort, “Attenuated Total Reflection. A New Principle for the Production of Useful Infra-Red Reflection Spectra of Organic Compounds,” *Spectrochimica Acta,* **17**, 698 (1961).

[519] R. H. Renard, “Total Reflection: A New Evaluation of the Goos-Hdnchen Shift,” *J. Opt. Soc. Am.*, **54**, 1190 (1964).

[520] N. J. Harrick, *Internal Reflection Spectroscopy*, Wiley, New York, 1967.

[521] P. W. Baumeister, “Optical Tunneling and its Application to Optical Filters,” *Appl. Opt.*, **6**, 897 (1967).

[522] H. K. V. Lotsch, “Reflection and Refraction of a Beam of Light at a Plane Interface,” *J. Opt. Soc. Am.*, **58**, 551 (1968).

[523] B. R, Horowitz and T. Tamir, “Lateral Displacement of a Light Beam at a Dielectric Interface,” *J. Opt. Soc. Am.*, **61**, 586 (1971).

[524] C. Imbert, “Calculation and Experimental Proof of the Transverse Shift Induced by Total Internal Reflection of a Circularly Polarized Light Beam,” *Phys. Rev.*, **D-5**, 787 (1972).

[525] A. W. Snyder and J. D. Love, “Goos-Hänchen shift,” *Appl. Opt.*, **15**, 236 (1976).

[526] M. McGuirk and C. K. Carniglia, “An angular spectrum representation approach to the Goos-Hänchen shift,” *J. Opt. Soc. Am.*, **67**, 103 (1977).

[527] J. J. Cowan and B. Anicin, “Longitudinal and transverse displacements of a bounded microwave beam at total internal reflection,” *J. Opt. Soc. Am.*, **67**, 1307 (1977).

[528] S. Kozaki and H. Sakurai, “Characteristics of a Gaussian beam at a dielectric interface,” *J. Opt. Soc. Am.*, **68**, 508 (1978).

[529] G. Müller, K. Abraham, and M. Schaldach, “Quantitative ATR spectroscopy: some basic considerations,” *Appl. Opt.*, **20**, 1182 (1981).

[530] I. R. Chandler, V. P. Tomaselli, and K. D. Moller, “Attenuated total reflection method for obtaining the optical constants of powders,” *Appl. Opt.*, **22**, 4099 (1983).

[531] S. Zhu, et al., “Frustrated Total Internal Reflection: A Demonstration and Review,” *Am. J. Phys.*, **54**, 601 (1986).

[532] H. M. Lai, F. C. Cheng, and W. K. Tang, “Goos-Hänchen effect around and off the critical angle,” *J. Opt. Soc. Am.*, **A-3**, 550 (1986).

[533] C. C. Chan and T. Tamir, “Beam phenomena at and near critical incidence upon a dielectric interface,” JOSA, **A-4**, 655 (1987).

[534] D. Gingell, O. S. Heavens, and J. S. Mellor, “General electromagnetic theory of total internal reflection fluorescence: the quantitative basis for mapping cell-substratum topography,” *J. Cell Sci.*, **87**, 677 (1987).

[535] O. S. Heavens, “Cell studies of total internal reflection fluorescence: effect of lipid membranes,” *J. Cell Sci.*, **95**, 175 (1990).

[536] R. C. Reddick, R. J. Warmack, and T. L. Ferrell, “New Form of Scanning Optical Microscopy,” *Phys. Rev.*, **B-39**, 767 (1989).

[537] J. Navasquillo, V. Such, and F. Pomer, “A general method for treating the incidence of a plane electromagnetic wave on a plane interface between dielectrics,” *Am. J. Phys.*, **57**, 1109 (1989).

[538] F. Pomer and J. Navasquillo, “The fields of a bounded electromagnetic beam propagating through an air gap between two dielectrics for frustrated total reflection,” *Am. J. Phys.*, **58**, 763 (1990).

[539] R. C. Reddick, et al., “Photon Scanning Tunneling Microscopy,” *Rev. Sci. Instrum.*, **61**, 3669 (1990).

[540] F. Albiol, S. Navas, and M. V. Andres, “Microwave Experiments on Electromagnetic Evanescent Waves and Tunneling Effect,” *Am. J. Phys.*, **61**, 165 (1993).

[541] S. Sainov, V. Sainov, and G. Stoilov, “Interferometer based on total internal reflection,” *Appl. Opt.*, **34**, 2848 (1995).

[542] A. Salari and R. E. Young, “Application of attenuated total reflectance FTIR spectroscopy to the analysis of mixtures of pharmaceutical polymorphs,” *Int. J. Pharmaceutics*, **163**, 157 (1998).

[543] F. de Fornel, *Evanescent Waves*, Springer-Verlag, Berlin, 2000.

[544] A. Haibel, G. Nimtz, and A. A. Stahlhofen, “Frustrated Total Reflection: The Double-Prism Revisited,” *Phys. Rev.*, **E-63**, 047601 (2001).

[545] L. Li and J. A. Dobrowolski, "High-Performance Thin-Film Polarizing Beam Splitter Operating at Angles Greater than the Critical Angle," *Appl. Opt.*, **39**, 2754 (2000).

[546] L. Li, "The Design of Optical Thin Film Coatings with Total and Frustrated Total Internal Reflection," *Opt. & Photon. News*, p. 24, September 2003.

[547] F. P. Zanella, et al., "Frustrated Total Internal Reflection: A Simple Application and Demonstration," *Am. J. Phys.*, **71**, 494 (2003).

[548] D. A. Papathanassoglou and B. Vohnsen, "Direct Visualization of Evanescent Optical Waves," *Am. J. Phys.*, **71**, 670 (2003).

[549] E. R. Van Keuren, "Refractive index measurement using total internal reflection," *Am. J. Phys.*, **73**, 611 (2005).

Edward Richard Van Keuren

[550] E. Marengo, "Monitoring of paintings under exposure to UV light by ATR-FT-IR spectroscopy and multivariate control charts," *Vibr. Spectrosc.*, **40**, 225 (2006).

[551] C-F. Li, "Unified theory for Goos-Hänchen and Imbert-Fedorov effects," *Phys. Rev.*, **A-76**, 013811 (2007).

[552] C-W. Chen, et al., "Optical temperature sensing based on the Goos-Hänchen effect," *Appl. Opt.*, **46**, 5347 (2007).

Surface Plasmons

[553] U. Fano, "The Theory of Anomalous Diffraction Gratings and of Quasi-Stationary Waves on Metallic Surfaces (Sommerfeld's Waves)," *J. Opt. Soc. Am.*, **31**, 213 (1941).

[554] R. H. Ritchie, "Plasma losses by fast electrons in thin films," *Phys. Rev.*, **106**, 874 (1957).

[555] E. Kretschmann and H. Raether, "Radiative decay of nonradiative surface plasmons excited by light," *Z. Naturforsch. A*, **23**, 2135 (1968).

[556] A. Otto, "Excitation of Nonradiative Surface Plasma Waves in Silver by the Method of Frustrated Total Reflection," *Z. Physik*, **216**, 398 (1968).

[557] E. N. Economou, "Surface Plasmons in Thin Films," *Phys. Rev.*, **182**, 539 (1969).

[558] E. Kretschmann, "Die Bestimmung optischer Konstanten von Metallen durch Anregung von Oberflächenplasmaschwingungen," *Z. Physik*, **241**, 313 (1971).

[559] A. S. Barker, "An Optical Demonstration of Surface Plasmons on Gold," *Am. J. Phys.*, **42**, 1123 (1974).

[560] M. Cardona, "Fresnel Reflection and Surface Plasmons," *Am. J. Phys.*, **39**, 1277 (1971).

[561] H. J. Simon, D. E. Mitchell, and J. G. Watson, "Surface Plasmons in Silver Films—A Novel Undergraduate Experiment," *Am. J. Phys.*, **43**, 630 (1975).

[562] J. D. Swalen, et al., "Plasmon surface polariton dispersion by direct optical observation," *Am. J. Phys.*, **48**, 669 (1980).

[563] W. P. Chen and J. M. Chen, "Use of Surface Plasma Waves for Determination of the Thickness and Optical Constants of Thin Metallic Films," *J. Opt. Soc. Am.*, **71**, 189 (1981).

[564] A. D. Boardman, ed., *Electromagnetic Surface Modes*, Wiley, New York, 1982.

[565] F. Abeles, "Surface Plasmon (SEW) Phenomena," in *Electromagnetic Surface Excitations*, R. F. Wallis and G. I. Stegeman, eds., Springer-Verlag, Berlin, 1986.

[566] I. J. Higgins and C. R. Lowe, "Introduction to the Principles and Applications of Biosensors," *Phil. Trans. Roy. Soc. Lond.*, **B-316**, 3 (1987).

[567] B. Rothenhäusler and W. Knoll "Surface-plasmon microscopy," *Nature*, **332**, 615 (1988).

[568] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag, Berlin, 1988.

[569] V. I. Baibakov, V. N. Datsko, and Y. V. Kistovich, "Experimental Discovery of Zenneck's Surface Electromagnetic Wave," *Sov. Phys. Usp.*, **32**, 378 (1989).

[570] K. Welford, "Surface Plasmon-Polaritons and their Uses," *Opt. Quant. Electr.*, **23**, 1 (1991).

[571] J. R. Sambles, G. W. Bradbery, and F. Yang, "Optical Excitation of Surface Plasmons: An Introduction," *Contemp. Phys.*, **32**, 173 (1991).

[572] B. Liedberg, C. Nylander, and I. Lundström, "Biosensing with surface plasmon resonance—how it all started," *Biosens. & Bielectron.*, **10**, no. 8, pp. i-ix, (1995).

[573] J. J. Ramsden, "Optical Biosensors," *J. Molec. Recognition*, **10**, 109 (1997).

[574] H. F. Ghaemi, et al., "Surface plasmons enhance optical transmission through subwavelength holes," *Phys. Rev.*, **B-58**, 6779 (1998).

[575] J. Homola, S. S. Yee, and G. Gauglitz, "Surface plasmon resonance sensors: review," *Sens. Actuators B*, **54**, 3 (1999).

[576] L. Martin-Moreno, et al., "Theory of extraordinary optical transmission through subwavelength hole arrays," *Phys. Rev. Lett.*, **86**, 1114 (2001). (2001).

[577] K. Kurihara, K. Nakamura, and K. Suzuki, "Asymmetric SPR Sensor Response Curve-Fitting Equation for the Accurate Determination of SPR Resonance Angle," *Sens. Actuators B*, **86**, 49 (2002).

[578] J. Homola, "Present and future of surface plasmon resonance biosensors," *Anal. Bioanal. Chem.*, **377**, 528 (2003).

[579] D. A. Schultz, "Plasmon resonant particles for biological detection," *Curr. Opin. Biotechnol.*, **14**, 13 (2003).

[580] B. Lee, "Review of the Present Status of Optical Fiber Sensors," *Opt. Fiber Technol.*, **9**, 57 (2003).

[581] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface Plasmon Subwavelength Optics," *Nature*, **424**, 824 (2003).

[582] T. M. Chinowsky, et al., "Performance of the Spreeta 2000 integrated surface 4 plasmon resonance affinity senso," *Sens. Actuators B*, **91**, 266 (2003).

[583] A. I. Csurgay and W. Porod, "Surface Plasmon Waves in Nanoelectronic Circuits," *Int. J. Circ. Th. Appl.*, **32**, 339 (2004).

[584] C. Girard, "Near Fields in Nanostructures," *Rep. Prog. Phys.*, **68**, 1883 (2005).

[585] A. Ramanavièius, et al., "Biomedical Application of Surface Plasmon Resonance Biosensors (Review)," *Acta Med. Lituanica*, **12**, 1 (2005).

[586] Y. P. Bliokh, et al., "Visualization of the complex refractive index of a conductor by frustrated total internal reflection," *Appl. Phys. Lett.*, **89**, 021908 (2006).

[587] P. Lecaruyer, et al., "Generalization of the Rouard method to an absorbing thin-film stack and application to surface plasmon resonance," *Appl. Opt.*, **45**, 8419 (2006).

[588] L. Novotny and B. Hecht, *Principles of Nano-Optics*, Cambridge Univ. Press, Cambridge, 2006.

[589] S. A. Maier, *Plasmonics: Fundamentals and Applications*, Springer, New York, 2007.

[590] H. A. Atwater, "The Promise of Plasmonics," *Sci. Amer.*, April 2007, p. 56.

[591] J. M. Pitarke, et al., "Theory of Surface Plasmons and Surface-Plasmon Polaritons," *Rep. Progr. Phys.*, **70**, 1 (2007).

Thin Films

[592] O. S. Heavens, *Optical Properties of Thin Solid Films*, Butterworths Scientific Publications, London, 1955, and Dover Publications, New York, 1991.

[593] A. Vasicek, *Optics of Thin Films*, North-Holland Publishing Co., Amsterdam, 1960.

[594] H. A. Macleod, *Thin-Film Optical Filters*, American Elsevier, New York, 1969.

[595] E. Hecht and A. Zajac, *Optics*, Addison-Wesley, Reading, MA, 1974.

[596] Z. Knittl, *Optics of Thin Films*, McGraw-Hill, New York, 1976.

[597] O. S. Heavens, *Thin Film Physics*, Methuen, London, 1970.

[598] M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, 1980.

[599] A. Thelen, *Design of Optical Interference Coatings*, McGraw-Hill, New York, 1989.

[600] J. A. Dobrowolski, "Optical Properties of Films and Coatings," in *Handbook of Optics, vol.I*, M. Bass, et al., eds., McGraw-Hill, New York, 1995.

[601] O. S. Heavens, J. Ring, and S. D. Smith, "Interference Filters for the Infrared," *Spectrochimica Acta*, **10**, 179 (1957).

[602] H. Van de Stadt and J. M. Muller, "Multimirror Fabry-Perot Interferometers," *J. Opt. Soc. Am.*, **A-2**, 1363 (1985).

[603] A. Zheng, J. S. Seely, R. Hunneman, and G. J. Hawkins, "Design of Narrowband Filters in the Infrared Region," *Infrared Phys.*, **31**, 237 (1991).

[604] J. M. Bennett, "Polarizers," in *Handbook of Optics, vol.II*, M. Bass, et al., eds., McGraw-Hill, New York, 1995. See also, J. M. Bennett, "Polarization," *ibid., vol.I.*

[605] H. Bach and D. Krause, eds., *Thin Films on Glass*, Springer-Verlag, Berlin, 1997.

[606] Lord Rayleigh, "On the Reflection of Light from a Regularly Stratified Medium," *Proc. R. Soc. London*, Ser. A, **93**, 565 (1917).

[607] M. Banning, "Practical Methods of Making and Using Multilayer Filters," *J. Opt. Soc. Am.*, **37**, 792 (1947).

[608] R. B. Muchmore, "Optimum Band Width for Two Layer Anti-Reflection Films," *J. Opt. Soc. Am.*, **38**, 20 (1948).

[609] F. Abelès, "Recherches sur la Propagation des Ondes Electromagnetiques Sinusoidales dans les Milieux Stratifiés," *Ann. Physique*, Ser.12, **5**, 596 (1950) and Part II, p.706.

[610] P. J. Leurgans, "The Impedance Concept in Thin Film Optics," *J. Opt. Soc. Am.*, **41**, 714 (1951).

[611] L. I. Epstein, "The Design of Optical Filters," *J. Opt. Soc. Am.*, **42**, 806 (1952).

[612] L. I. Epstein, "Improvements in Heat-Reflecting Filters," *J. Opt. Soc. Am.*, **45**, 360 (1952).

[613] P. T. Scharf, "Transmission Color in Camera Lenses,", *J. SMPTE*, **59**, 191 (1952).

[614] W. Weinstein, "Computations in Thin Film Optics," *Vacuum*, **4**, 3 (1954).

[615] F. E. Carlson, et al., "Temperature Reduction in Motion-Picture and Television Studios Using Heat-Control Coatings,", *J. SMPTE*, **65**, 136 (1956).

[616] H. H. Schroeder and A. F. Turner, "A Commercial Cold Reflector," *J. SMPTE*, **69**, 351 (1960).

[617] L. Young, "Synthesis of Multiple Antireflection Films Over a Prescribed Frequency Band," *J. Opt. Soc. Am.*, **51**, 967 (1961).

[618] P. H. Berning, "Use of Equivalent Films in the Design of Infrared Multilayer Antireflection Coatings," *J. Opt. Soc. Am.*, **52**, 431 (1962).

[619] L. A. Catalan, "Some Computed Optical Properties of Antireflection Coatings," *J. Opt. Soc. Am.*, **52**, 437 (1962).

[620] J. Cox, G. Hass, and A. Thelen, "Triple-Layer Antireflection Coatings on Glass for the Visible and Near Infrared," *J. Opt. Soc. Am.*, **52**, 965 (1962).

[621] F. Abelès, "Methods for Determining Optical Parameters of Thin Films," in *Progress in Optics*, vol. II, E. Wolf, ed., North-Holland Publishing Co., Amsterdam, 1963.

[622] A. Thelen, "Multilayer Filters with Wide Transmittance Bands," *J. Opt. Soc. Am.*, **53**, 1266 (1963).

[623] J. J. Vera, "Some Properties of Multi-Layer Films with Periodic Structure," *Opt. Acta*, **11**, 315 (1964).

[624] A. Thelen, "Equivalent Layers in Multilayer Filters," *J. Opt. Soc. Am.*, **56**, 1533 (1966).

[625] J. Arndt and P. Baumeister, "Reflectance and Phase Envelopes of an Iterated Multilayer," *J. Opt. Soc. Am.*, **56**, 1760 (1966).

[626] A. F. Turner and P. W. Baumeister, "Multilayer Mirrors with High Reflectance Over an Extended Spectral Region," *Appl. Opt.*, **5**, 69 (1966).

[627] L. Young, "Multilayer Interference Filters with Narrow Stop Bands," *Appl. Opt.*, **6**, 297 (1967).

[628] H. F. Mahlein and G. Schollmeier, "Analysis and Synthesis of Periodic Optical Resonant Reflectors," *Appl. Opt.*, **8**, 1197 (1969).

[629] P. B. Clapham, M. J. Downs, and R. J. King, "Some Applications of Thin Films to Polarization Devices," *Appl. Opt.*, **8**, 1965 (1969).

[630] E. Delano and R. J. Pegis, "Methods of Synthesis for Dielectric Multilayer Filters," in *Progress in Optics*, vol. VII, E. Wolf, ed., North-Holland Publishing Co., Amsterdam, 1969.

[631] A. Thetford, "A Method of Designing Three-Layer Anti-Reflection Coatings," *Opt. Acta*, **16**, 37 (1969).

[632] A. Musset and A. Thelen, "Multilayer Antireflection Coatings," in *Progress in Optics*, vol. VIII, E. Wolf, ed., North-Holland Publishing Co., Amsterdam, 1970.

[633] P. Baumeister and G. Pincus, "Optical Interference Coatings," *Sci. Amer.*, **223**, 59, Dec. 1970.

[634] L. Young and E. G. Cristal, "Low-Pass and High-Pass Filters Consisting of Multilayer Dielectric Stacks," *IEEE Trans. Microwave Theory Tech.*, **MTT-14**, 75 (1966).

[635] A. Thelen, "Design of Optical Minus Filters," *J. Opt. Soc. Am.*, **61**, 365 (1971).

[636] A. Kucirkova, "Synthesis of Multiple Anti-Reflection Films by the Rational Function Method," *Opt. Acta*, **18**, 577 (1971).

[637] J. A. Dobrowolski and S. H. C. Piotrowski, "Refractive Index as a Variable in the Numerical Design of Optical Thin Film Systems," *Appl. Opt.*, **21**, 1502 (1982).

[638] P. H. Berning, "Principles of Design of Architectural Coatings," *Appl. Opt.*, **22**, 4127 (1983).

[639] R. Swanepoel, "Determination of the Thickness and Optical Constants of Amorphous Silicon," *J. Phys. E: Sci. Instrum.*, **16**, 1214 (1983).

[640] J. S. Seeley, "Simple Nonpolarizing High-Pass Filter," *Appl. Opt.*, **24**, 742 (1985).

[641] R. Herrmann, "Quarterwave Layers: Simulation by Three Thin Layers of Two Materials," *Appl. Opt.*, **24**, 1183 (1985).

[642] P. Baumeister, "Antireflection Coatings with Chebyshev or Butterworth Response: Design,", *Appl. Opt.*, **25**, 4568 (1986).

[643] A. V. Tikhonravov, "Some Theoretical Aspects of Thin-Film Optics and their Applications," *Appl. Opt.*, **32**, 5417 (1993).

[644] J. Mouchart, J. Begel, and E. Duda, "Modified MacNeille Cube Polarizer for a Wide Angular Field," *Appl. Opt.*, **28**, 2847 (1989).

[645] E. Cojocaru, "Comparison of Theoretical Performances for Different Single-Wavelength Thin-Film Polarizers," *Appl. Opt.*, **31**, 4501 (1992).

[646] L. Li and J. A. Dobrowolski, "Visible Broadband, Wide-Angle, Thin-Film Multilayer Polarizing Beam Splitter," *Appl. Opt.*, **35**, 2221 (1996).

[647] K. V. Popov, J. A. Dobrowolski, A. V. Tikhonravov, and B. T. Sullivan, "Broadband High-Reflection Multilayer Coatings at Oblique Angles of Incidence," *Appl. Opt.*, **36**, 2139 (1997).

[648] P. W. Baumeister, "Rudiments of the Design of an Immersed Polarizing Beam Divider with Narrow Spectral Bandwidth and Enhanced Angular Acceptance," *Appl. Opt.*, **36**, 3610 (1997).

[649] J. Ciosek, J. A. Dobrowolski, G. A. Clarke, and G. Laframboise, "Design and Manufacture of All-Dielectric Nonpolarizing Beam Splitters," *Appl. Opt.*, **38**, 1244 (1999).

[650] L. Li and J. A. Dobrowolski, "High-Performance Thin-Film Polarizing Beam Splitter Operating at Angles Greater than the Critical Angle," *Appl. Opt.*, **39**, 2754 (2000).

[651] A. Thelen, "Design Strategies for Thin Film Optical Coatings,", in Ref. [605].

[652] B. Danielzik, M. Heming, D. Krause, and A. Thelen, "Thin Films on Glass: An Established Technology,", in Ref. [605].

[653] W. T. Doyle, "Scattering Approach to Fresnel's Equations and Brewster's Law," *Am. J. Phys.*, **53**, 463 (1985).

[654] K. Sato, et al., "Measurement of the Complex Refractive Index of Concrete at 57.5 GHz," *IEEE Trans. Antennas Propagat.*, **AP-44**, 35 (1996).

[655] R. K. Zia, "Symmetric Fresnel Equations: An Energy Conservation Approach," *Am. J. Phys.*, **56**, 555 (1988).

[656] O. S. Heavens and H. M. Liddell, "Staggered Broad-Band Reflecting Multilayers," *Appl. Opt.*, **5**, 373 (1966).

[657] G. N. Henderson, T. K. Gaylord, and E. N. Glytsis, "Ballistic Electron Transport in Semiconductor Heterostructures and Its Analogies in Electromagnetic Propagation in General Dielectrics," *Proc. IEEE*, **79**, 1643 (1991).

Birefringent Multilayer Films

[658] M. F. Weber, C. A. Stover, L. R. Gilbert, T. J. Nevitt, and A. J. Ouderkirk, "Giant Birefringent Optics in Multilayer Polymer Mirrors," *Science*, **287**, 2451 (2000).

[659] H. Schopper, "Zur Optik dünner dopplebrechender und dichroitischer Schichten," *Z. Physik*, **132**, 146 (1952).

[660] D. A. Holmes, "Exact Theory of Retardation Plates," *J. Opt. Soc. Am.*, **54**, 1115 (1964).

[661] D. A. Holmes and D. L. Feucht, "Electromagnetic Wave Propagation in Birefringent Multilayers," *J. Opt. Soc. Am.*, **56**, 1763 (1966).

[662] L. P. Mosteller and F. Wooten, “Optical Properties and Reflectance of Uniaxial Absorbing Crystals,” *J. Opt. Soc. Am.*, **58**, 511 (1968).

[663] R. H. W. Graves, “Determination of Optical Constants of Anisotropic Crystals,” *J. Opt. Soc. Am.*, **59**, 1225 (1969).

[664] A. Wünsche, “Neue Formeln für die Reflexion und Brechung des Lichtes an anisotropen Medien,” *Ann. Physik*, series 7, **25**, 201 (1970).

[665] S. Teitler and B. W. Henvis, “Refraction in Stratified Anisotropic Media,” *J. Opt. Soc. Am.*, **60**, 830 (1970).

[666] A. S. Marathay, “Matrix-Operator Description of the Propagation of Polarized Light through Cholesteric Liquid Crystals,” *J. Opt. Soc. Am.*, **60**, 1363 (1970).

[667] C. Altman and S. G. Lipson, “Reciprocity Relations in Light Propagation through a Multilayer Birefringent System,” *J. Opt. Soc. Am.*, **61**, 1460 (1971).

[668] D. den Engelsen, “Ellipsometry of Anisotropic Films,” *J. Opt. Soc. Am.*, **61**, 1460 (1971).

[669] D. W. Berreman, “Optics in Stratified and Anisotropic Media: 4 × 4 Matrix Formulation,” *J. Opt. Soc. Am.*, **62**, 502 (1972).

[670] J. Schesser and G. Eichman, “Propagation of Plane Waves in Biaxially Anisotropic Layered Media,” *J. Opt. Soc. Am.*, **62**, 786 (1972).

[671] D. J. De Smet, “Ellipsometry of Anisotropic Thin Films,” *J. Opt. Soc. Am.*, **64**, 631 (1974).

[672] D. J. De Smet, “Generalized Ellipsometry and the 4 × 4 Matrix Formalism,” *Surface Science*, **56**, 293 (1976).

[673] J. J. Stamnes and G. C. Sherman, “Reflection and Refraction of an Arbitrary Wave at a Plane Interface Separating Two Uniaxial Crystals,” *J. Opt. Soc. Am.*, **67**, 683 (1977).

[674] P. Yeh, “Electromagnetic Propagation in Birefringent Layered Media,” *J. Opt. Soc. Am.*, **69**, 742 (1979).

[675] P. Yeh, “Optics of Anisotropic Layered Media: A New 4 × 4 Matrix Algebra,” *Surface Science*, **96**, 41 (1980).

[676] R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland Publishing Company, Amsterdam, 1977.

[677] I. J. Hodgkinson and Q. H. Wu, *Birefringent Thin Films and Polarizing Elements*, World Scientific, Singapore, 1997.

[678] I. J. Hodgkinson, S. Kassam, and Q. H. Wu, “Eigenequations and Compact Algorithms for Bulk and Layered Anistropic Optical Media: Reflection and Refraction at a Crystal-Crystal Interface,” *J. Comput. Phys.*, **133**, 75 (1997).

Chiral Media

[679] A. Lakhtakia, ed., *Selected Papers on Natural Optical Activity*, SPIE Milestone Series, vol. MS 15, SPIE Optical Engineering Press, Bellingham, WA, 1990.

[680] M. P. Silverman, *Waves and Grains*, Princeton University Press, Princeton, NJ, 1998.

[681] E. U. Condon, “Theories of Optical Rotatory Power,” *Rev. Mod. Phys.*, **9**, 432 (1937). Reprinted in [679].

[682] F. I. Fedorov, “On the Theory of Optical Activity in Crystals,” *Optics and Spectroscopy*, **6**, 49 (1959), and *ibid.*, p.237, and with B. V. Bokut, p.342. Reprinted in [679].

[683] B. D. H. Tellegen, “The Gyrator, A New Electrical Network Element,” *Philips Res. Reports*, **3**, 81 (1948). Reprinted in [679].

[684] M. P. Silverman and R. B. Sohn, “Effects of Circular Birefringence on Light Propagation and Reflection,” *Am. J. Phys.*, **54**, 69 (1986).

[685] M. P. Silverman, “Reflection and Refraction at the Surface of a Chiral Medium: Comparison of Gyrotropic Constitutive Relations Invariant or Noninvariant Under a Duality Transformation,” *J. Opt. Soc. Am.*, **A-3**, 830 (1986).

[686] A. Lakhtakia, V. V. Varadan, and V. K. Varadan, “A Parametric Study of Microwave Reflection Characteristics of a Planar Achiral-Chiral Interface,” *IEEE Trans. Electrom. Compat.*, **EMC-28**, 90 (1986).

[687] A. Lakhtakia, V. V. Varadan, and V. K. Varadan, “Field Equations, Huygens’s Principle, Integral Equations, and Theorems for Radiation and Scattering of Electromagnetic Waves in Isotropic Chiral Media,” *J. Opt. Soc. Am.*, **A-5**, 175 (1988).

[688] S. Bassiri, C. H. Papas, and N. Engheta, “Electromagnetic Wave Propagation Through and Dielectric-Chiral Interface and Through an Chiral Slab,” *J. Opt. Soc. Am.*, **A-5**, 1450 (1988).

[689] J. C. Monson, “Radiation and Scattering in Homogeneous General Biisotropic Regions,” *IEEE Trans. Antennas Propagat.*, **AP-38**, 227 (1990). Reprinted in [679].

[690] D. L. Jaggard, A. R. Mickelson, and C. H. Papas, “On Electromagnetic Waves in Chiral Media,” *Appl. Phys.*, **18**, 211 (1979).

[691] D. L. Jaggard and X. Sun, “Theory of Chiral Multilayers,” *J. Opt. Soc. Am.*, **A-9**, 804 (1992).

[692] K. M. Flood and D. L. Jaggard, “Band-Gap Structure for Periodic Chiral Media,” *J. Opt. Soc. Am.*, **A-13**, 1395 (1996).

[693] S. F. Mason, “Form Pasteur to Parity Nonconservation: Theories of the Origin of Molecular Chirality,” in Ref. [694].

[694] N. Berova, K. Nakanishi, and R. W. Woody, eds., *Circular Dichroism*, 2nd ed., Wiley-VCH, New York, 2000.

[695] A. Lakhtakia and W. S. Weiglhofer, “Are Linear, Nonreciprocal, Biisotropic Media Forbidden?,” *IEEE Trans. Microwave Theory Tech.*, **MTT-42**, 1715 (1994).

Gyrotropic Media

[696] E. V. Appleton, “Wireless Studies of the Ionosphere,” *J. IEE*, **71**, 642 (1932).

[697] D. R. Hartree, “The Propagation of Electromagnetic Waves in a Refracting Medium in a Magnetic Field,” *Proc. Camb. Phil. Soc.*, **27**, 143 (1931).

[698] R. S. Lawrence, C. G. Little, and H. J. A. Chivers, “A Survey of Ionospheric Effects Upon Earth-Space Radio Propagation,” *Proc. IEEE*, **52**, 4 (1964).

[699] K. Davies, *Ionospheric Radio Waves*, Blaisdell Publishing Co., Waltham, MA, 1969.

[700] D. G. Swanson, *Plasma Waves*, Academic Press, San Diego, CA, 1989.

[701] F. Bloch, “Nuclear Induction,” *Phys. Rev.*, **70**, 460 (1946).

[702] F. Bloch, W. W. Hansen, and M. Packard, “The Nuclear Induction Experiment,” *Phys. Rev.*, **70**, 474 (1946).

[703] D. Polder, “On the Theory of Ferromagnetic Resonance,” *Philos. Mag.*, **40**, 99 (1949).

[704] D. Polder, “On the Phenomenology of Ferromagnetic Resonance,” *Phys. Rev.*, **73**, 1120 (1948).

[705] W. A. Yager, J. K. Galt, F. R. Merritt, and E. A. Wood, “Ferromagnetic Resonance in Nickel Ferrite,” *Phys. Rev.*, **80**, 744 (1950).

[706] C. L. Hogan, “The Ferromagnetic Faraday Effect at Microwave Frequencies and its Applications, The Microwave Gyrator” *Bell Syst. Tech. J.*, **31**, 1 (1952).

[707] C. L. Hogan, “The Ferromagnetic Faraday Effect at Microwave Frequencies and its Applications,” *Rev. Mod. Phys.*, **25**, 253 (1953).

[708] M. T. Weiss and A. G. Fox, “Magnetic Double Refraction at Microwave Frequencies,” *Phys. Rev.*, **88**, 146 (1952).

[709] N. Bloembergen, “Magnetic Resonance in Ferrites,” *Proc. IRE*, **44**, 1259 (1956).

[710] B. Lax and K. J. Button, *Microwave Ferrites and Ferrimagnetics*, McGraw Hill, New York, 1962.

[711] D. M. Bolle and L. Lewin, “On the Definitions of Parameters in Ferrite-Electromagnetic Wave Interactions,” *IEEE Trans. Microwave Theory Tech.*, **MTT-21**, 118 (1974).

[712] K. Button, “Microwave Ferrite Devices: The First Ten Years,” *IEEE Trans. Microwave Theory Tech.*, **MTT-32**, 1088 (1984).

Photonic and Other Bandgaps

[713] L. Brillouin, *Wave Propagation in Periodic Structures*, Dover, New York, 1953.

[714] C. Elachi, “Waves in Active and Passive Periodic Structures: A Review,” *Proc. IEEE*, **64**, 1666 (1976).

[715] P. Yeh, A. Yariv, and C-S. Hong, "Electromagnetic Propagation in Periodic Stratified Media. I. General Theory," *J. Opt. Soc. Am.*, **67**, 423 (1977), and "II. Birefringence, Phase Matching, and X-Ray Lasers," *ibid.*, p. 438.

[716] A. Yariv and P. Yeh, *Optical Waves in Crystals: Propagation and Control of Laser Radiation*, Wiley, New York, 1984.

[717] P. Yeh, *Optical Waves in Layered Media*, Wiley, New York, 1988.

[718] D. W. L. Sprung and H. Wu, "Scattering by a Finite Periodic Potential," *Am. J. Phys.*, **61**, 1118 (1993).

[719] J. L. Rosner, "Reflectionless Approximations to Potentials with Band Structure," *Ann. Phys.*, **200**, 101 (1990).

[720] Photonic Band Gap Bibliography. See web site Ref. [1320].

[721] E. Yablonovitch, "Photonic Band-Gap Structure,", *J. Opt. Soc. Am.*, **B-10**, 283 (1992).

[722] E. Yablonovitch, "Photonic Crystals,", *J. Mod. Opt.*, **41**, 173 (1994)

[723] J. B. Pendry, "Photonic Band Structures," *J. Mod. Opt.*, **41**, 209 (1994).

[724] P. St. J. Russel, "Photonic Band Gaps," *Phys. World*, **5**, 37, August 1992.

[725] R. D. Meade, et al., "Novel Applications of Photonic Band Gap Materials: Low-Loss Bends and High Q Cavities," *J. Appl. Phys.*, **75**, 4753 (1994).

[726] J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*, Princeton University Press, Princeton, NJ, 1995.

[727] C. M. Soukoulis, "Photonic Band Gap Materials: The "Semiconductors" of the Future?," *Physica Scripta*, **T66**, 146(1996).

[728] J. D. Joannopoulos, P. R. Villeneuve, and S. Fan, "Photonic Crystals: Putting an New Twist on Light," *Nature*, **386**, 143 (1997).

[729] M. Jacoby, "Photonic Crystals: Whole Lotta Holes," *Chem & Eng. News*, November 23, 1998, p. 38.

[730] Special Issue on Electromagnetic Crystal Structures, Design, Synthesis, and Applications, *J. Lightwave Technol.*, **17**, no. 11, November 1999.

[731] Mini-Special Issue on Electromagnetic Crystal Structures, Design, Synthesis, and Applications, *IEEE Trans. Microwave Theory Tech.*, **MTT-47**, no. 11, November 1999.

[732] T. A. Birks, J. C. Knight, and P. St., J. Russell, "Endlessly Single-Mode Photonic Crystal Fiber," *Optics Lett.*, **22**, 961 (1997).

[733] O. Painter, R. K. Lee, A. Scherer, A. Yariv, J. D. O'Brien, P. D. Dapkus, and I. Kim, "Two-Dimensional Photonic Band-Gap Defect Mode Laser," *Science*, **284**, 1819 (1999).

[734] O. J. Painter, A. Husain, A. Scherer, J. D. O'Brien, I. Kim, and P. D. Dapkus, "Room Temperature Photonic Crystal Defect Lasers at Near-Infrared Wavelengths in InGaAsP," *J. Lightwave Technol.*, **17**, 2082 (1999).

[735] Y. Fink, D. J. Ripin, S. Fan, C. Chen, J. D. Joannopoulos, and E. L. Thomas, "Guiding Optical Light in Air Using an All-Dielectric Structure," *J. Lightwave Technol.*, **17**, 2039 (1999).

[736] Y. Fink, J. N. Winn, S. Fan, C. Chen, J. Michel, J. D. Joannopoulos, and E. L. Thomas, "A Dielectric Omnidirectional Reflector," *Science*, **282**, 1679 (1998).

[737] D. N. Chigrin, A. V. Lavrinenko, D. A. Yarotsky, and S. V. Gaponenko, "Observation of Total Omnidirectional Reflection from a One-Dimensional Dielectric Lattice," *Appl. Phys. A*, **68**, 25 (1999).

[738] J. N. Winn, Y. Finn, S. Fan, and J. D. Joannopoulos, "Omnidirectional Reflection from a One-Dimensional Photonic Crystal," *Opt. Lett.*, **23**, 1573 (1998).

[739] D. N. Chigrin, A. V. Lavrinenko, D. A. Yarotsky, and S. V. Gaponenko, "All-Dielectric One-Dimensional Periodic Structures for Total Omnidirectional Reflection and Partial Spontaneous Emission Control," *J. Lightwave Technol.*, **17**, 2018 (1999).

[740] P. St. J. Russell, S. Tredwell, and P. J. Roberts, "Full Photonic Bandgaps and Spontaneous Emission Control in 1D Multilayer Dielectric Structures,"*Optics Commun.*, **160**, 66 (1999).

[741] J. P. Dowling, "Mirror on the Wall: You're Omnidirectional After All?" *Science*, **282**, 1841 (1998).

[742] D. Normile, "Cages for Light Go from Concept to Reality," *Science*, **286**, 1500 (1999).

[743] J. M. Bendickson and J. P. Dowling, "Analytic Expressions for the Electromagnetic Mode Density in Finite, One-Dimensional, Photonic Band-Gap Structures," *Phys. Rev. E*, **53**, 4107 (1996).

[744] J. P. Dowling, "Parity, Time-Reversal and Group Delay for Inhomogeneous Dielectric Slabs: Application to Pulse Propagation in Finite, One-Dimensional, Photonic Bandgap Structures," *IEE Proc. J, Optoelectron.*, **145**, 420 (1998).

[745] J. P. Dowling, "Dipole Emission in Finite Photonic Bandgap Structures: An Exactly Solvable One-Dimensional Model," *J. Lightwave Technol.*, **17**, 2142 (1999).

[746] `www.sspectra.com/designs/omnirefl.html`, Software Spectra, Inc., "Dielectric Omnidirectional Reflector," February 1999.

[747] H. Kogelnik and C. V. Shank, "Coupled-Wave Theory of Distributed Feedback Lasers," *J. Appl. Phys.*, **43**, 2327 (1972).

[748] H. A. Haus, "Grating-Filter Transformation Chart," *Electron. Lett.*, **11**, 553 (1975).

[749] H. Kogelnik, "Filter Response of Nonuniform Almost-Periodic Structures," *Bell Syst. Tech. J.*, **55**, 109 (1976).

[750] L. A. Weller-Brophy and D. G. Hall, "Analysis of Waveguide Gratings: Application of Rouard's Method," *J. Opt. Soc. Am.*, **A-2**, 863 (1985).

[751] L. A. Weller-Brophy and D. G. Hall, "Analysis of Waveguide Gratings: A Comparison of the Results of Rouard's Method and Coupled-Mode Theory," *J. Opt. Soc. Am.*, **A-4**, 60 (1987).

[752] M. Yamada and K. Sakuda, "Analysis of Almost-Periodic Distributed Feedback Slab Waveguides via a Fundamental Matrix Approach," *Appl. Opt.*, **26**, 3474 (1987).

[753] K. A. Winick, "Effective-Index Method and Coupled-Mode Theory for Almost-Periodic Waveguide Gratings: A Comparison," *Appl. Opt.*, **31**, 757 (1992).

[754] N. Matuschek, F. X. Kärtner, and U. Keller, "Exact Coupled-Mode Theories for Multilayer Interference Coatings with Arbitrary Strong Index Modulations," *IEEE J. Quant. Electr.*, **QE-33**, 295 (1997).

[755] T. Erdogan, "Fiber Grating Spectra," *J. Lightwave Technol.*, **15**, 1277 (1997).

[756] C. R. Giles, "Lightwave Applications of Fiber Bragg Gratings," *J. Lightwave Technol.*, **15**, 1391 (1997).

[757] H. A. Haus and C. V. Shank, "Antisymmetric Taper of Distributed Feedback Lasers," *IEEE J. Quant. Electr.*, **QE-12**, 532 (1976).

[758] K. Utaka, S. Akiba, K. Sakai, and Y. Matsushima, "Analysis of Quarter-Wave-Shifted DFB Laser," *Electron. Lett.*, **20**, 326 (1984).

[759] K. Utaka, S. Akiba, K. Sakai, and Y. Matsushima, "$\lambda/4$-Shifted InGaAsP/InP DFB Lasers by Simultaneous Holographic Exposure of Positive and Negative Photoresists," *Electron. Lett.*, **20**, 326 (1984).

[760] R. C. Alferness, C. H. Joyner, M. D. Divino, M. J. R. Martyak, and L. L. Buhl, "Narrowband Grating Resonator Filters in InGaAsP/InP Waveguides," *Appl. Phys. Lett.*, **49**, 125 (1986).

[761] H. A. Haus and Y. Lai, "Theory of Cascaded Quarter Wave Shifted Distributed Feedback Resonators," *IEEE J. Quant. Electr.*, **QE-28**, 205 (1992).

[762] G. P. Agrawal and S. Radic, "Phase-Shifted Fiber Bragg Gratings and their Application for Wavelength Demultiplexing," *IEEE Photon. Technol. Lett.*, **6**, 995 (1994).

[763] R. Zengerle and O. Leminger, "Phase-Shifted Bragg-Grating Filters with Improved Transmission Characteristics," *J. Lightwave Technol.*, **13**, 2354 (1995).

[764] L. Wei and W. Y. Lit, "Phase-Shifted Bragg Grating Filters with Symmetrical Structures," *J. Lightwave Technol.*, **15**, 1405 (1997).

[765] F. Bakhti and P. Sansonetti, "Design and Realization of Multiple Quarter-Wave Phase-Shifts UV-Written Bandpass Filters in Optical Fibers," *J. Lightwave Technol.*, **15**, 1433 (1997).

[766] R. Kashyap, *Fiber Bragg Gratings*, Academic Press, San Diego, CA, 1999.

[767] S. V. Kartalopoulos, *Introduction to DWDM Technology*, IEEE Press, New York, 2000.

[768] S. S. Mester and H. Benaroya, "A Review of Periodic and Near-Periodic Structures," *Shock and Vibration*, **2**, 69 (1995).

[769] D. J. Mead, "Wave Propagation in Continuous Periodic Structures," *J. Sound and Vibration*, **190**, 495 (1996).

[770] R. S. Langley, N. S. Bardell, and P. M. Loasby, "The Optimal Design of Near-Periodic Structures to Minimize Vibration Transmission and Stress Levels," *J. Sound and Vibration*, **207**, 627 (1997).

[771] R. Martinez-Sala, et al., "Sound Attenuation by Sculpture," *Nature*, **378**, 241 (1995).

[772] M. M. Sigalas and E. N. Economou, "Attenuation of Multiple-Scattered Sound," *Europhys. Lett.*, **36**, 241 (1996).

[773] W. M. Robertson and J. F. Rudy, “Measurement of Acoustic Stop Bands in Two-Dimensional Periodic Scattering Arrays,” *J. Acoust. Soc. Am.*, **104**, 694 (1998).

[774] M. S. Kushwaha and Djafari-Rouhani, “Sonic Stop-Bands for Periodic Arrays of Metallic Rods: Honeycomb Structure,” *J. Sound and Vibration*, **218**, 697 (1998).

[775] C. Rubio, et al., “The Existence of Full Gaps and Deaf Bands in Two-Dimensional Sonic Crystals,” *J. Lightwave Technol.*, **17**, 2202 (1999).

[776] Z. Liu, X. Zhang, Y. Mao, Y. Y. Zhu, Z. Yang, C. T. Chang, and P. Sheng, “Locally Resonant Sonic Materials,” *Science*, **289**, 1734 (2000).

Fiber-Optic Filters

[777] B. Moslehi, J. W. Goodman, M. Tur, and H. J. Shaw, “Fiber-Optic Lattice Signal Processing,” *Proc. IEEE*, **72**, 909 (1984).

[778] K. Jackson, S. Newton, B. Moslehi, M. Tur, C. Cutler, J. W. Goodman, and H. J. Shaw, “Optical Fiber Delay-Line Signal Processing,” *IEEE Trans. Microwave Theory Tech.*, **MTT-33**, 193 (1985).

[779] E. M. Dowling and D. L. MacFarlane, “Lightwave Lattice Filters for Optically Multiplexed Communication Systems,” *J. Lightwave Technol.*, **12**, 471 (1994).

[780] C. Madsen and J. Zhao, *Optical Filter Design and Analysis: A Signal Processing Approach*, Wiley, New York, 1999.

Quarter-Wave Transformers

[781] R. E. Collin, “Theory and Design of Wide-Band Multisection Quarter-Wave Transformers,” *Proc. IRE*, **43**, 179 (1955).

[782] R. E. Collin and J. Brown, “The Design of Quarter-Wave Matching Layers for Dielectric Surfaces,” *Proc. IEE*, **103C**, 153 (1955).

[783] S. B. Cohn, “Optimum Design of Stepped Transmission-Line Transformers,” *IRE Trans. Microwave Theory Tech.*, **MTT-3**, 16 (1955).

[784] H. J. Riblet, “General Synthesis of Quarter-Wave Impedance Transformers,” *IRE Trans. Microwave Theory Tech.*, **MTT-5**, 36 (1957).

[785] R. Levy, “A Guide to the Practical Application of Chebyshev Functions to the Design of Microwave Components,” *Proc. IEE*, **106C**, 193 (1959).

[786] L. Young, “Stepped-Impedance Transformers and Filter Prototypes,” *IEEE Trans. Microwave Theory Tech.*, **MTT-10**, 339 (1962).

[787] L. Young, “Unit Real Functions in Transmission-Line Circuit Theory,” *IEEE Trans. Circuit Th.*, **CT-7**, 247 (1960).

[788] R. Levy, “Tables of Element Values for the Distributed Low-Pass Prototype Filter,” *IEEE Trans. Microwave Theory Tech.*, **MTT-13**, 514 (1965).

[789] C. S. Gledhill and A. M. H. Issa, “On Synthesis of Particular Unit Real Functions of Reflection Coefficient,” *IEEE Trans. Microwave Theory Tech.*, **MTT-17**, 57 (1969).

[790] C. S. Gledhill and A. M. H. Issa, “Exact Solutions of Stepped Impedance Transformers Having Maximally Flat and Chebyshev Characteristics,” *IEEE Trans. Microwave Theory Tech.*, **MTT-17**, 379 (1969).

[791] G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance Matching Networks, and Coupling Structures*, Artech House, Dedham, MA, 1980.

Linear Prediction, Speech, Geophysics, Network, and Function Theory

[792] S. J. Orfanidis, *Optimum Signal Processing*, 2nd ed., McGraw-Hill, New York, 1988.

[793] E. A. Robinson and S. Treitel, “Maximum Entropy and the Relationship of the Partial Autocorrelation to the Reflection Coefficients of a Layered System,” *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-28**, 22 (1980).

[794] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, Upper Saddle River, NJ, 1978.

[795] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.

[796] J. A. Ware and K. Aki, “Continuous and Discrete Inverse Scattering Problems in a Stratified Elastic Medium, I,” *J. Acoust. Soc. Am.*, **45**, 91 (1969).

[797] H. Wakita, “Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms,” *IEEE Trans. Audio Electroacoust.*, **AU-21**, 417 (1973).

[798] B. S. Atal and S. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the Speech Wave,” *J. Acoust. Soc. Am.*, **50**, 637 (1971).

[799] E. A. Robinson and S. Treitel, “Digital Signal Processing in Geophysics,” in *Applications of Digital Signal Processing*, A. V. Oppenheim, ed., Prentice Hall, Upper Saddle River, NJ, 1978.

[800] E. A. Robinson and S. Treitel, *Geophysical Signal Analysis*, Prentice Hall, Upper Saddle River, NJ, 1980.

[801] J. F. Claerbout, *Fundamentals of Geophysical Data Processing*, McGraw-Hill, New York, 1976.

[802] J. F. Claerbout, “Synthesis of a Layered Medium from its Acoustic Transmission Response,” *Geophysics*, **33**, 264 (1968).

[803] F. Koehler and M. T. Taner, “Direct and Inverse Problems Relating Reflection Coefficients and Reflection Response for Horizontally Layered Media,” *Geophysics*, **42**, 1199 (1977).

[804] J. M. Mendel and F. Habibi-Ashrafi, “A Survey of Approaches to Solving Inverse Problems for Lossless Layered Media Systems,” *IEEE Trans. Geosci. Electron.*, **GE-18**, 320 (1980).

[805] E. A. Robinson, “A Historical Perspective of Spectrum Estimation,” *Proc. IEEE*, **70**, 885 (1982).

[806] E. A. Robinson, “A Spectral Approach to Geophysical Inversion by Lorentz, Fourier, and Radon Transforms,” *Proc. IEEE*, **70**, 1039 (1982).

[807] K. P. Bube and R. Burridge, “The One-Dimensional Problem of Reflection Seismology,” *SIAM Rev.*, **25**, 497 (1983).

[808] J. G. Berryman and R. R. Green, “Discrete Inverse Methods for Elastic Waves in Layered Media,” *Geophysics*, **45**, 213 (1980).

[809] K. M. Case, “Inverse Scattering, Orthogonal Polynomials, and Linear Estimation,” in *Topics in Functional Analysis, Advances in Mathematics Supplementary Studies*, vol.3, I. C. Gohberg and M. Kac, eds., Academic Press, New York, 1978.

[810] F. J. Dyson, “Old and New Approaches to the Inverse Scattering Problem,” in *Studies in Mathematical Physics*, E. H. Lieb, B. Simon, and A. S. Wightman, eds., Princeton University Press, Princeton, NJ, 1976.

[811] R. G. Newton, “Inversion of Reflection Data for Layered Media: A Review of Exact Methods,” *Geophys. J. R. Astron. Soc.*, **65**, 191 (1981).

[812] O. Brune, “Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency,”, *J. Math. and Phys.*, **10**, 191 (1931).

[813] P. I. Richards, “A Special Class of Functions With Positive Real Part in a Half Plane,” *Duke Math. J.*, **14**, 777 (1947).

[814] H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York (1945).

[815] M. E. Van Valkenburg, *Introduction to Modern Network Synthesis*, Wiley, New York, 1960.

[816] H. J. Carlin and A. B. Giordano, *Network Theory, An Introduction to Reciprocal and Nonreciprocal Circuits*, Prentice Hall, Upper Saddle River, NJ, (1964).

[817] R. M. Fano, “Theoretical Limitations on the Broadband Matching of Arbitrary Impedances,” *J. Franklin Inst.*, **249**, 57 and 139 (1950).

[818] W-K. Chen, *Broadband Matching*, World Scientific, Singapore, 1988.

[819] I. Schur, “On Power Series which are Bounded in the Interior of the Unit Circle, I and II,” in *I. Schur Methods in Operator Theory and Signal Processing, Operator Theory: Advances and Applications*, vol.18, I. Gohberg, ed., Birkhäuser, Boston, 1986.

[820] T. Kailath, “A Theorem of I. Schur and its Impact on Modern Signal Processing,”, *ibid.*

[821] T. Kailath, A. M. Bruckstein, and D. Morgan, “Fast Matrix Factorization via Discrete Transmission Lines,”, *Lin. Alg. Appl.*, **75**, 1 (1985).

[822] A. Yagle and B. C. Levy, “The Schur Algorithm and its Applications,” *Acta Applic. Math.*, **3**, 255 (1985).

[823] P. P. Vaidyanathan and S. K. Mitra, "Discrete Version of Richard's Theorem and Applications to Cascaded Lattice Realization of Digital Filter Transfer Functions," *IEEE Trans. Circ. Syst.*, **CAS-33**, 26 (1986).

[824] R. J. Duffin, "Algorithms for Classical Stability Problems," *SIAM Rev.*, **11**, 196 (1969).

[825] A. J. Berkhout, "Stability and Least-Squares Estimation," *Automatica*, **11**, 637 (1975).

[826] P. P. Vaidyanathan and S. K. Mitra, "A Unified Structural Interpretation of Some Well-Known Stability-Test Procedures for Linear Systems," *Proc. IEEE*, **75**, 478 (1987).

Geometrical Optics and Ray Tracing

[827] S. Cornbleet, "Geometrical Optics Reviewed: A New Light on an Old Subject," *Proc. IEEE*, **71**, 471 (1983).

[828] S. Cornbleet, *Microwave and Geometrical Optics*, Academic Press, London, 1994.

[829] G. W. Forbes, "On Variational Problems in Parametric Form," *Am. J. Phys.*, **59**, 1130 (1991).

[830] E. G. Rawson, D. R. Herriott, and J. McKenna, "Analysis of Refractive Index Distributions in Cylindrical, Graded-Index Glass Rods (GRIN Rods) Used as Image Relays," *Appl. Opt.*, **9**, 753 (1970).

[831] E. W. Marchand, "Ray Tracing in Gradient-Index Media," *J. Opt. Soc. Am.*, **60**, 1 (1970).

[832] W. Streifer and K. B. Paxton, "Analytic Solution of Ray Equations in Cylindrically Inhomogeneous Guiding Media. 1: Meridional Rays," *Appl. Opt.*, **10**, 769 (1971).

[833] K. B. Paxton and W. Streifer, "Analytic Solution of Ray Equations in Cylindrically Inhomogeneous Guiding Media. Part 2: Skew Rays," *Appl. Opt.*, **10**, 1164 (1971).

[834] E. W. Marchand, "Ray Tracing in Cylindrical Gradient-Index Media," *Appl. Opt.*, **11**, 1104 (1972).

[835] R. Guenther, *Modern Optics*, Wiley, New York, 1990.

[836] ITU Recommendation **ITU-R P.453-6**, (1997), "The Radio Refractive Index: Its Formula and Refractivity Data." Available from [1315].

[837] ITU Recommendation **ITU-R P.834-2**, (1997), "Effects of Tropospheric Refraction on Radiowave Propagation." Available from [1315].

Waveguides and Transmission Lines

[838] N. Marcuvitz, ed., *Waveguide Handbook*, Dover Publications, New York, 1965.

[839] T. Saad, *Microwave Engineer's Handbook*, vols.I and II, Artech House, Dedham, MA, 1971.

[840] J. A. Staniforth, *Microwave Transmission*, Wiley, New York, 1972.

[841] K. Chang, ed., *Handbook of Microwave and Optical Components*, vol.1, Wiley, New York, 1989.

[842] R. E. Collin, *Field Theory of Guided Waves*, 2nd ed., IEEE Press, Piscataway, NJ, 1991.

[843] A. W. Lines, G. R. Nicoll, and A. M. Woodward, "Some Properties of Waveguides with Periodic Structure," *Proc. IEE*, **97**, Pt.III, 263 (1950).

[844] A. F. Harvey, "Periodic and Guiding Structures at Microwave Frequencies," *IEEE Trans. Microwave Theory Tech.*, **MTT-8**, 30 (1960).

[845] A. F. Harvey, *Microwave Engineering*, Academic Press, London, 1963.

[846] R. E. Collin, *Foundations of Microwave Engineering*, McGraw-Hill, New York, 1966.

[847] R. S. Elliott, *An Introduction to Guided Waves and Microwave Circuits*, Prentice Hall, Upper Saddle River, NJ, 1993.

[848] D. Marcuse, *Light Transmission Optics*, 2nd ed., Van Nostrand Reinhold, New York, 1982.

[849] A. Yariv, *Optical Electronics*, 3d ed., Holt, Rinehart, and Winston, Inc., New York, 1985.

[850] G. P. Agrawal, *Nonlinear Fiber Optics*, 3/e, Academic, New York, 2001.

[851] A. Hasegawa, *Optical Solitons in Fibers*, Springer-Verlag, Berlin, 1989.

[852] P. Diament, *Wave Transmission and Fiber Optics*, Macmillan, New York, 1990.

[853] H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice Hall, Upper Saddle River, NJ, 1984.

[854] D. L. Lee, *Electromagnetic Principles of Integrated Optics*, Wiley, New York, 1986.

[855] C. S. Lee, S. W. Lee, and S. L. Chuang, "Plot of Modal Field Distribution in Rectangular and Circular Waveguides", *IEEE Trans. Microwave Theory Tech.*, **MTT-33**, 271 (1985).

[856] D. J. White, "Adding Plane Waves to Find the Complete TM and TE Wave Solutions for Metallic Rectangular Waveguides," *Am. J. Phys.*, **51**, 1115 (1983).

[857] J. F. Lotspeich, "Explicit General Eigenvalue Solutions for Dielectric Slab Waveguides," *Appl. Opt.*, **14**, 327 (1975).

[858] S. Rosenstark, *Transmission Lines in Computer Engineering*, McGraw-Hill, New York, 1994.

[859] R. E. Matick, *Transmission Lines for Digital and Communication Networks*, IEEE Press, New York, 1995.

[860] J. C. Freeman, *Fundamentals of Microwave Transmission Lines*, Wiley, New York, 1996.

[861] "The Bergeron Method," Texas Instruments Application Report, SDYA014, October 1996, available online from `www.ti.com/sc/docs/psheets/abstract/apps/sdya014.htm`.

[862] E. O. Hammerstad and O. Jensen, "Accurate Models for Microstrip Computer-Aided Design," *IEEE MTT-S Digest International Microwave Symposium*, p.408, 1980, reprinted in Ref. [866].

[863] H. Wheeler, "Transmission Line Properties of Parallel Strips Separated by a Dielectric Sheet," *IEEE Trans. Microwave Theory Tech.*, **MTT-13**, 172 (1965).

[864] I. J. Bahl and D. K. Trivedi, "A Designer's Guide to Microstrip Line," *Microwaves*, **16**, May 1977, p. 174.

[865] I. J. Bahl and R. Garg, "A Designer's Guide to Stripline Circuits," *Microwaves*, **17**, January 1978, p. 90.

[866] T. Itoh, Ed., *Planar Transmission Line Structures*, IEEE Press, New York, 1987

[867] E. H. Fooks and R. A. Zakarevicius, *Microwave Engineering Using Microstrip Circuits*, Prentice Hall, Upper Saddle River, NJ, 1989.

[868] F. Gardiol, *Microstrip Circuits*, Wiley, New York, 1994.

[869] H. M. Barlow and A. L. Cullen, "Surface Waves," *Proc. IEE*, Pt.III, **100**, 329 (1953).

[870] S. A. Schelkunoff, "Anatomy of Surface Waves," *IRE Trans. Antennas Propagat.*, **AP-7**, 133 (1959).

[871] T. Tamir and A. A. Oliner, "Guided Complex Waves, Part 1" *Proc. IEE*, **110**, 310 (1963), and "Part 2,", *ibid.*, p. 325.

[872] A. Hessel, "General Characteristics of Traveling-Wave Antennas," in Ref. [9], vol.2, p. 151.

[873] T. Tamir, "Leaky-Wave Antennas," in Ref. [9], vol.2, p. 259.

[874] F. J. Zucker, "Surface-Wave Antennas," in Ref. [9], vol.2, p. 298.

[875] T. Tamir and F. Y. Koo, "Varieties of Leaky Waves and Their Excitation Along Multilayer Structures," *IEEE J. Quant. Electr.*, **QE-22**, 544 (1986).

[876] J. Zenneck, "Über die Fortpflanzung ebener electromagnetische Wellen längs einer ebenen Leiterfläche und ihre Beziehung zur drachtlosen Telegraphie," *Annalen der Physik*, **23**, 846 (1907).

[877] J. Fan, A. Dogariu, and L. J. Wang, "Amplified Total Internal Reflection," *Opt. Express*, **11**, 299 (2003).

[878] A. E. Siegman, "Evanescent Gain: Does It really Exist?," Stanford University, EE Department Preprint, unpublished, 2003.

[879] P. K. Tien, "Integrated Optics and New Wave Phenomena in Optical Waveguides," *Rev. Mod. Phys.*, **49**, 361 (1977).

Coupled Transmission Lines and Crosstalk

[880] S. A. Schelkunoff and T. M. Odarenko, "Crosstalk Between Coaxial Transmission Lines," *Bell Syst. Tech. J.*, **16**, 144 (1937).

[881] S. O. Rice, "Steady-State Solutions of Transmission Line Equations," *Bell Syst. Tech. J.*, **20**, 131 (1941).

[882] D. B. Jarvis, "The Effects of Interconnections on High-Speed Logic Circuits," *IEEE Trans. Electron. Comput.*, **EC-12**, 476 (1963).

[883] H. Amemiya, "Time-Domain Analysis of Multiple Parallel Transmission Lines," *RCA Rev.*, **28**, 241 (1967).

[884] J. A. DeFalco, "Reflection and Crosstalk in Logic Circuit Interconnections," *IEEE Spectrum*, **7**, July 1977, p.44.

[885] J. C. Isaacs and N. A. Strakhov, “Crosstalk in Uniformly Coupled Lossy Transmission Lines,” *Bell Syst. Tech. J.*, **52**, 101 (1973).

[886] A. Deutsch, et al., “High-Speed Signal Propagation on Lossy Transmission Lines,” *IBM J. Res. Dev.*, **34**, 601 (1990).

[887] C. R. Paul, “Literal Solutions for Time-Domain Crosstalk on Lossless Transmission Lines,” *IEEE Trans. Electromagn. Compat.*, **EMC-34**, 433 (1992).

[888] H. W. Johnson and M. Graham, *High-Speed Digital design*, Prentice Hall, Upper Saddle River, NJ, 1993.

[889] J. A. Brandão Faria, *Multiconductor Transmission-Line Structures*, Wiley, New York, 1993.

[890] C. R. Paul, *Analysis of Multiconductor Transmission Lines*, Wiley, New York, 1994.

[891] C. R. Paul, “Decoupling the Multiconductor Transmission Line Equations,”, *IEEE Trans. Microwave Theory Tech.*, **MTT-44**, 1429 (1996).

[892] J-F. Mao, O. Wing, and F-Y. Chang, “Synthesis of Coupled Transmission Lines,” *IEEE Trans. Circ. Syst., I*, **44**, 327 (1997).

[893] A. Deutsch, et al., “When are Transmission-Line Effects Important for On-Chip Interconnections?,” *IEEE Trans. Microwave Theory Tech.*, **MTT-45**, 1836 (1997).

[894] A. Deutsch, “Electrical Characteristics of Interconnections for High-Performance Systems,” *Proc. IEEE*, **86**, 315 (1998).

[895] T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*, Wiley, New York, 2000.

[896] J. A. Davis, et al., “Interconnect Limits on Gigascale Integration (GSI) in the 21st Century,” *Proc. IEEE*, **89**, 305 (2001).

[897] C. R. Paul, “Solution of the Transmission-Line Equations Under the Weak-Coupling Assumption,” *IEEE Trans. Electromagn. Compat.*, **EMC-44**, 413 (2002).

Coupled-Mode Theory

[898] J. R. Pierce, “Coupling of Modes of Propagation,” *J. Appl. Phys.*, **25**, 179 (1954).

[899] S. E. Miller, “Coupled Wave Theory and Waveguide Applications,” *Bell Syst. Tech. J.*, **33**, 661 (1954).

[900] J. S. Cook, “Tapered Velocity Couplers,” *Bell Syst. Tech. J.*, **34**, 807 (1955).

[901] W. H. Louisell, “Analysis of the Single Tapered Mode Coupler,” *Bell Syst. Tech. J.*, **34**, 853 (1955).

[902] W. H. Louisell, *Coupled Mode and Parametric Electronics*, Wiley, New York, 1960.

[903] C. W. Barnes, “Conservative Coupling Between Modes of Propagation—A Tabular Summary,” *Proc. IEEE*, **52**, 64 (1964).

[904] H. Kogelnik, “Coupled Wave Theory for Thick Hologram Gratings,” *Bell Syst. Tech. J.*, **48**, 2909 (1969).

[905] D. Marcuse, “The Coupling of Degenerate Modes in Two Parallel Dielectric Waveguides,” *Bell Syst. Tech. J.*, **50**, 1791 (1971).

[906] A. W. Snyder, “Coupled-Mode Theory for Optical Fibers,” *J. Opt. Soc. Am.*, **62**, 1267 (1972).

[907] A. Yariv, “Coupled-Mode Theory for Guided-Wave Optics,” *IEEEJ, Quant. Electron.*, **QE-9**, 919 (1973).

[908] H. F. Taylor and A. Yariv, “Guided Wave Optics,” *Proc. IEEE*, **62**, 1044 (1974).

[909] H. Kogelnik, “Theory of Dielectric Waveguides,” in *Integrated Optics*, T. Tamir, ed., Springer-Verlag, New York, 1975.

[910] E. Marom, O. G. Ramer, and S. Ruschin, “Relation Between Normal-Mode and Coupled-Mode Analyses of Parallel Waveguides,” *IEEEJ, Quant. Electron.*, **QE-20**, 1311 (1984).

[911] A. Hardy and W. Streifer, “Coupled Mode Solutions of Multiwaveguide Systems,” *IEEEJ, Quant. Electron.*, **QE-22**, 528 (1986).

[912] A. W. Snyder, Y. Chen, and A. Ankiewicz, “Coupled Waves on Optical Fibers by Power Conservation,” *J. Lightwave Technol.*, **7**, 1400 (1989).

[913] H. A, Haus and W-P. Huang, “Coupled-Mode Theory,” *Proc. IEEE*, **79**, 1505 (1991).

[914] W-P. Huang, “Coupled-Mode Theory for Optical Waveguides: An Overview,” *J. Opt. Soc. Am.*, **A-11**, 963 (1994).

[915] W-P. Yuen, “On the Different Formulations of the Coupled-Mode Theory for Parallel Dielectric Waveguides,” *J. Lightwave Technol.*, **12**, 82 (1994).

[916] R. März, *Integrated Optics*, Artech House, Boston, 1995.

[917] B. Little, “Filter Synthesis for Coupled Waveguides,” *IEEEJ, Quant. Electron.*, **QE-15**, 1149 (1997).

[918] N. Matuschek, G. Steinmeyer, and U. Keller, “Relation Between Coupled-Mode Theory and Equivalent Layers for Multilayer Interference Coatings,” *Appl. Opt.*, **39**, 1626 (2000).

[919] M. McCall, “On the Application of Coupled Mode Theory for Modeling Fiber Bragg Gratings,” *J. Lightwave Technol.*, **18**, 236 (2000).

Diffuse Reflection and Transmission

[920] A. Schuster, “Radiation Through a Foggy Atmosphere,” *Astroph. J.*, **21**, 1 (1905). Reprinted in [922] and [923].

[921] F. Kottler, “The Elements of Radiative Transfer,” *Progress in Optics*, E. Wolf, ed., vol. III, North-Holland Publishing Co., Amsterdam, 1964.

[922] D. H. Menzel, ed., *Selected Papers on the Transfer of Radiation*, Dover Publications, New York, 1966.

[923] C. F. Bohren, ed., *Selected Papers on Scattering in the Atmosphere*, SPIE Milestone Series, vol. MS-7, SPIE Press, Bellingham, WA, 1989.

[924] C. F. Bohren, “Multiple Scattering of Light and Some of its Observable Consequences,” *Am. J. Phys.*, **55**, 524 (1987).

[925] P. Kubelka and F. Munk, “Ein Beitrag zur Optik der Farbanstriche,” *Zeit. Tech. Physik*, **12**, 593 (1931). Translated by S. Westin in `www.graphics.cornell.edu/~westin/pubs/kubelka.pdf`.

[926] P. Kubelka, “New Contributions to the Optics of Intensely Light-Scattering Materials, Part I,” *J. Opt. Soc. Am.*, **38**, 448 (1948).

[927] P. Kubelka, “New Contributions to the Optics of Intensely Light-Scattering Materials, Part II: Nonhomogeneous Layers,” *J. Opt. Soc. Am.*, **44**, 330 (1954).

[928] D. B. Judd, *Color in Business, Science, and Industry*, 2nd ed., Wiley, New York, 1963.

[929] W. W. Wedlandt and H. G. Hecht, *Reflectance Spectroscopy*, Interscience, Wiley, New York, 1966.

[930] G. Kortum, *Reflectance Spectroscopy*, Springer Verlag, New York, 1969.

[931] W. G. Egan and T. W. Hilgeman, *Optical Properties of Inhomogeneous Materials*, Academic Press, New York, 1979.

[932] S. Wan, R. R. Anderson, and J. A. Parrish, “Analytical Modeling for the Optical Properties of the Skin with In Vitro and In Vivo Applications,” *Photochem. Photobiol.*, **34**, 493 (1981).

[933] B. Chaudhuri and S. C. Som, “Experimental Verification of the Matrix Formulation of the Schuster-Kubelka-Munk Theory of Diffusing Layers,” *J. Optics*, **12**, 245 (1981).

[934] R. Molenaar, J. J. ten Bosch, and J. R. Zijp, “Determination of Kubelka-Munk Scattering and Absorption Coefficients by Diffuse Illumination,” *Appl. Opt.*, **38**, 2068 (1999).

[935] V. I. Haltrin, “Diffuse Reflection Coefficient of a Stratified Sea,” *Appl. Opt.*, **38**, 932 (1999).

[936] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, IEEE Press, Piscataway, NJ, 1997.

Impedance Matching

[937] N. Balabanian, “Impedance Matching,” *IEEE Trans. Microwave Theory Tech.*, **MTT-3**, 53 (1955).

[938] M. A. Hamid and M. M. Yunik, “On the Design of Stepped Transmission-Line Transformers,” *IEEE Trans. Microwave Theory Tech.*, **MTT-15**, 528 (1967).

[939] G. N. French and E. H. Fooks, “The Design of Stepped Transmission-Line Transformers,” *IEEE Trans. Microwave Theory Tech.*, **MTT-16**, 885 (1968).

[940] R. M. Arnold, “Transmission Line Impedance Matching Using the Smith Chart,” *IEEE Trans. Microwave Theory Tech.*, **MTT-21**, 977 (1974).

[941] J. H. Lepoff, “Matching: When Are Two Lines Better Than One?,” *Microwaves*, March 1981, p.74.

[942] G. N. French and E. H. Fooks, “Double Section Matching Transformers,” *IEEE Trans. Microwave Theory Tech.*, **MTT-17**, 719 (1969).

[943] F. Regier, “Series-Section Transmission Line Impedance Matching”, *QST*, July 1978, p.14.

[944] M. W. Maxwell, “Another Look at Reflections,”, Parts 1–4, *QST*, April 1973 p.35, June, p.20, August, p.36, October, p.22.

[945] F. Witt, “Match Bandwidth of Resonant Antenna Systems,” *QST*, October 1991, p.21. See also, A. S. Griffith, “Match Bandwidth Revisited,” *QST*, June 1992, p.71.

[946] D. K. Belcher, “RF Matching Techniques, Design and Example,”, *QST*, October 1972, p.

[947] T. Dorbuck, “Matching-Network Design,”, *QST*, March 1979, p.24.

[948] J. S. Chakmanian, “Control VSWR Bandwidth in T-Section Networks,” *Microwaves*, July 1981, p.87.

[949] E. Wingfield, “New and Improved Formulas for the Design of Pi and Pi-L Networks,” *QST*, August 1983, p.23, and *QST*, January 1984, p.49.

[950] W. N. Caron, *Antenna Impedance Matching*, America Radio Relay League, Newington, CT, 1999.

[951] Hewlett Packard, “Interactive Impedance Matching Model,” available from the web site [1323].

[952] Y. L. Chow and K. L. Wan, “A Transformer of One-Third Wavelength in Two Sections – For a Frequency and Its First Harmonic,” *IEEE Microwave Wireless Compon. Lett.*, **12**, 22 (2002).

[953] C. Monzon, “Analytical Derivation of a Two-Section Impedance Transformer for a Frequency and Its First Harmonic,” *IEEE Microwave Wireless Compon. Lett.*, **12**, 381 (2002).

[954] C. Monzon, “A Small Dual-Frequency Transformer in Two Sections,” *IEEE Trans. Microwave Theory Tech.*, **51**, 1157 (2003).

[955] S. J. Orfanidis, “A Two-Section Dual-Band Chebyshev Impedance Transformer,” *IEEE Microwave Wireless Compon. Lett.*, **13**, 382 (2003).

S-Parameters

[956] H. Carlin, “The Scattering Matrix in Network Theory,” *IEEE Trans. Circ. Th.*, **CT-3**, 88 (1956).

[957] V. Belevitch, “Elementary Applications of the Scattering Formalism in Network Design,” *IEEE Trans. Circ. Th.*, **CT-3**, 97 (1956).

[958] D. C. Youla, “On Scattering Matrices Normalized to Complex Port Numbers,” *Proc. IRE*, **49**, 1221 (1961).

[959] D. C. Youla and P. M. Paterno, “Realizable Limits of Error for Dissipationless Attenuators in Mismatched Systems,” *IEEE Trans. Microwave Theory Tech.*, **MTT-12**, 289 (1964).

[960] K. Kurokawa, “Power Waves and the Scattering Matrix,” *IEEE Trans. Microwave Theory Tech.*, **MTT-13**, 194 (1965).

[961] R. F. Bauer and P. Penfield, Jr., “De-Embedding and Unterminating,” *IEEE Trans. Microwave Theory Tech.*, **MTT-22**, 282 (1974).

[962] G. E. Bodway, “Two-Port Power Flow Analysis Using Generalized Scattering Parameters,” *Microwave J.*, **10**, 61, May 1967.

[963] J. K. Hunton, “Analysis of Microwave Measurement Techniques by Means of Signal Flow Graphs,” *IEEE Trans. Microwave Theory Tech.*, **MTT-8**, 206 (1960).

[964] N. Kuhn, “Simplified Signal Flow Graph Analysis,” *Microwave J.*, Nov. 1963, p.61.

[965] F. Weinert, “Scattering Parameters Speed Design of High-Frequency Transistor Circuits,” *Electronics*, September 5, 1966, p.78.

[966] W. H. Froehner, “Quick Amplifier Design with Scattering Parameters,” *Electronics*, **40**, no.21, p.100, October 16, 1967.

[967] J. M. Rollett, “Stability and Power-Gain Invariants of Linear Twoports,” IRE Trans. Circuit Theory, **CT-9**, 29 (1962).

[968] D. Woods, “Reappraisal of the Unconditional Stability Criteria for Active 2-Port Networks in terms of S Parameters,” IEEE Trans. Circ. Syst., **CAS-23**, 73 (1976).

[969] R. P. Meys, “Review and Discussion of Stability Criteria for Linear 2-Ports,” *IEEE Trans. Circ. Syst.*, **CS-37**, 1450 (1990).

[970] M. L. Edwards and J. H. Sinsky, “A New Criterion for Linear 2-Port Stability Using a Single Geometrically Derived Parameter,” *IEEE Trans. Microwave Theory Tech.*, **MTT-40**, 2303 (1992).

[971] R. W. Anderson, “S-Parameter Design for Faster, More Accurate Network Design,” Hewlett Packard, Test & Measurement Application Note 95-1, available from the web site [1323].

[972] Hewlett Packard Application Note 1287-1, “Understanding the Fundamental Principles of Vector Network Analysis,” available from the web site [1323].

[973] H. Rothe and W. Dahlke, “Theory of Noisy Fourpoles,” *Proc. IRE*, **44**, 811 (1956). Reprinted in [983].

[974] R. W. Beatty, “Insertion Loss Concepts,” *Proc. IEEE*, **52**, 663 (1964).

[975] D. M. Kerns, “Definitions of V, I, Z, Y, a, b, G_{app}, and S,” *Proc. IEEE*, **55**, 892 (1967).

[976] P. Penfield, Jr., “Wave Representation of Amplifier Noise,” IRE Trans. Circuit Theory, **CT-9**, 84 (1962). Reprinted in [983].

[977] I. A. Harris, “Dependence of Receiver Noise-Temperature Measurement on Source Impedance,” *Electr. Lett.*, **2**, 130 (1966).

[978] H. Fukui, “Available Power Gain, Noise Figure, and Noise Measure of Two-Ports and their Graphical Representations,”, IEEE Trans. Circuit Theory, **CT-13**, 137 (1966).

[979] R. S. Tucker, “Low-Noise Design of Microwave Transistor Amplifiers,” *IEEE Trans. Microwave Theory Tech.*, **MTT-23**, 697 (1975).

[980] R. P. Meys, “A Wave Approach to the Noise Properties of Linear Microwave Devices,” *IEEE Trans. Microwave Theory Tech.*, **MTT-26**, 34 (1978).

[981] S. Withington, “Scattered Noise Waves in Microwave and mm-Wave Networks,” *Microwave J.*, **32**, 169, June 1989.

[982] C. Bowick, *RF Circuit Design*, Newness, Boston, 1980.

[983] H. Fukui, ed., *Low-Noise Microwave Transistors & Amplifiers*, IEEE Press, Piscataway, NJ 1981.

[984] T. T. Ha, *Solid-State Microwave Amplifier Design*, Wiley, New York, 1981.

[985] P. L. Abrie, *The Design of Impedance-Matching Networks for Radio-Frequency and Microwave Amplifiers*, Artech House, Dedham, MA, 1985.

[986] S. Y. Liao, *Microwave Circuit Analysis and Amplifier Design*, Prentice Hall, Upper Saddle River, NJ, 1987.

[987] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley, New York, 1990.

[988] M. W. Medley, *Microwaves and RF Circuits*, Artech House, Norwood, MA, 1992.

[989] G. Gonzalez, *Microwave Transistor Amplifiers*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1997.

[990] D. M. Pozar, *Microwave Engineering*, 2nd ed., Wiley, New York, 1998.

[991] “High-Frequency Transistor Primer,”, Parts I–IV, in Ref. [1325].

[992] “Maximizing Accuracy in Noise Figure Measurements,” Hewlett-Packard Product Note 85719A-1, (1992).

[993] D. Vondran, “Noise Figure Measurement: Corrections Related to Match and Gain,” *Microwave J.*, **42**, 22, March 1999.

[994] D. Boyd, “Calculate the Uncertainty of NF Measurements,” *Microwave J.*, **42**, 93, October 1999.

[995] “Noise Figure Measurement Accuracy—The Y-Factor Method,” Agilent Technologies, Appl. Note 57-2, (2001).

Transmitting, Receiving, and Scattering Properties of Antennas

[996] G. Sinclair, “The Transmission and Reception of Elliptically Polarized Waves,” *Proc. IRE*, **38**, 148 (1950).

[997] C. T. Tai, “On the Definition of the Effective Aperture of Antennas,” *IEEE Trans. Antennas Propagat.*, **AP-9**, 224 (1961).

[998] H. T. Friis, “A Note on a Simple Transmission Formula,” *Proc. IRE*, **34**, 254 (1946).

[999] H. T. Friis, “Introduction to Radio and Radio Antennas,”, *IEEE Spectrum*, **8**, 55, April 1971.

[1000] R. T. Bush, “The Antenna Formula: An Application of Single-Slit Diffraction Theory,” *Am. J. Phys.*, **55**, 350 (1987).

[1001] D. C. Hogg, “Fun with Friis Free-Space Transmission Formula,” *IEEE Antennas and Propagation Mag.*, **35**, 33, August 1993.

[1002] R. E. Collin, "The Receiving Antenna," in Ref. [9], part 1.

[1003] M. R. Andrews, P. P. Mitra, and R. deCarvalho, "Tripling the Capacity of Wireless Communications Using Electromagnetic Polarization," *Nature*, **409**, 316 (2001). See also, H. L. Bertoni, "Talk Is Cheap in the City," *Nature*, **409**, 291 (2001).

[1004] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, eds., *Principles of Microwave Circuits*, McGraw-Hill., New York, 1947.

[1005] A. F. Stevenson, "Relations Between the Transmitting and Receiving Properties of Antennas," *Quart. Appl. Math.*, **5**, 369 (1948).

[1006] S. H. Dike and D. D. King, "The Absorption Gain and Back-Scattering Cross Section of the Cylindrical Antenna," *Proc. IRE*, **40**, 853 (1952).

[1007] Y-Y. Hu, "Back-Scattering Cross Section of a Center-Loaded Cylindrical Antenna," *IRE Trans. Antennas Propagat.*, **AP-6**, 140 (1958).

[1008] J. Brown, "A Generalized Form of the Aerial Reciprocity Theorem," *Proc. IEE*, **103C**, 472 (1958).

[1009] D. M. Kerns and E. S. Dayhoff, "Theory of Diffraction in Microwave Interferometry," *J. Res. Nat. Bur. Stand.*, **64B**, 1 (1960).

[1010] D. Midgley, "A Theory of Receiving Aerials Applied to the Reradiation of an Electromagnetic Horn," *Proc. IEE*, **108B**, 645 (1961).

[1011] R. F. Harrington, "Theory of Loaded Scatterers," *Proc. IEE*, **111**, 617 (1964).

[1012] R. J. Garbacz, "Modal Expansions for Resonance Scattering Phenomena," *Proc. IEEE*, **53**, 856 (1965).

[1013] J. K. Schindler, R. B. Mack, and P. Blacksmith, Jr., "The Control of Electromagnetic Scattering by Impedance Loading," *Proc. IEEE*, **53**, 993 (1965).

[1014] W. H. Kahn and H. Kurss, "Minimum Scattering Antennas," *IEEE Trans. Antennas Propagat.*, **AP-13**, 671 (1965).

[1015] R. B. Green, "Scattering from Conjugate-Matched Antennas," *IEEE Trans. Antennas Propagat.*, **AP-14**, 17 (1968).

[1016] A. T. De Hoop, "A Reciprocity Relation Between the Transmitting and the Receiving Properties of an Antenna," *Appl. Sci. Res.*, **19**, 90 (1968).

[1017] A. C. Gately, D. J. R. Stock, and B. R-S Cheo, "A Network Description for Antenna Problems," *Proc. IEEE*, **56**, 1181 (1968).

[1018] W. Wasylkiwsky and W. H. Kahn, "Theory of Mutual Coupling Among Minimum-Scattering Antennas," *IEEE Trans. Antennas Propagat.*, **AP-18**, 204 (1970).

[1019] W. Wasylkiwsky and W. H. Kahn, "Scattering Properties and Mutual Coupling of Antennas with Prescribed Radiation Pattern," *IEEE Trans. Antennas Propagat.*, **AP-18**, 741 (1970).

[1020] A. T. De Hoop and G. De Jong, "Power Reciprocity in Antenna Theory," *Proc. IEE*, **121**, 1051 (1974).

[1021] A. T. De Hoop, "The N-Port Receiving Antenna and its Equivalent Electrical Network," *Philips Res. Reports*, **30**, 302 (1975).

[1022] D. M. Kerns, *Plane-Wave Scattering Matrix Theory of Antennas and Antenna-Antenna Interactions*, Nat. Bur. Stand. Monograph 162, Washington 1981. See also, *J. Res. Nat. Bur. Stand.*, **80B**, 5 (1975).

[1023] P, G, Rogers, "Application of the Minimum Scattering Antenna Theory to Mismatched Antennas," *IEEE Trans. Antennas Propagat.*, **AP-34**, 1223 (1986).

[1024] R. M. Bevensee, "A Lower Bound to the Broad-Band Power Scattered from an Electrically Linear Antenna with a General Lumped Load," *IEEE Trans. Antennas Propagat.*, **AP-37**, 555 (1989).

[1025] R. C. Hansen, "Relationships Between Antennas as Scatterers and as Radiators," *Proc. IEEE*, **77**, 659 (1989).

[1026] J. Van Bladel, "On the Equivalent Circuit of a Receiving Antenna," *IEEE Antennas and Propagation Mag.*, **44**, no.1, p.164, February 2002.

[1027] A. W. Love, "Comment: On the Equivalent Circuit of Receiving Antenna," *IEEE Antennas and Propagation Mag.*, **44**, no.5, p.124, October 2002.

[1028] R. C. Johnson, "Absorption of Energy Incident Upon a Receiving Antenna," *Microwave J.*, **15**, no.12, p.35 December 1972.

[1029] R. E. Collin, "Limitations of The Thévenin and Norton Equivalent Circuits for a Receiving Antenna," *IEEE Antennas and Propagation Mag.*, **45**, no.2, p.119, April 2003. See also, *ibid.*, no.4, p.99, August 2003.

[1030] A. W. Love, "Comment on "Limitations of The Thévenin and Norton Equivalent Circuits for a Receiving Antenna"," *IEEE Antennas and Propagation Mag.*, **45**, no.4, p.98, August 2003.

[1031] J. Bach Andersen and R. G. Vaughan, "Transmitting, Receiving, and Scattering Properties of Antennas," *IEEE Antennas and Propagation Mag.*, **45**, no.4, p.93, August 2003.

Noise Temperature

[1032] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, **32**, 97 (1928).

[1033] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, **32**, 110 (1928).

[1034] W. R. Bennett, *Electrical Noise*, McGraw-Hill, New York, 1960.

[1035] B. M. Oliver, "Thermal and Quantum Noise," *Proc. IEEE*, **53**, 436 (1965).

[1036] J. B. Johnson, "Electronic Noise: The First Two Decades," *IEEE Spectrum*, **8**, 42, Feb. 1971.

[1037] H. T. Friis, "Noise Figures of Radio Receivers," *Proc. IRE*, **32**, 419 (1944). Reprinted in [983].

[1038] J. R. Pierce, "Physical Sources of Noise," *Proc. IRE*, **44**, 601 (1956).

[1039] A. E. Siegman, "Thermal Noise in Microwave Systems, Part 1,", *Microwave J.*, **4**, p.81, March 1961, see also, "Part 2,", *ibid.*, p.66, April 1961, and "Part 3," *ibid.*, p.93, May 1961.

[1040] J. S. Wells, W. C. Daywitt, and C. K. S. Miller, "Measurement of Effective Temperature of Microwave Noise Sources," *IEEE Trans. Instr. Meas.*, **IM-13**, 17 (1964).

[1041] B. L. Seidel and C. T. Stelzried, "A Radiometric Method for Measuring the Insertion Loss of Radome Materials," *IEEE Trans. Microwave Theory Tech.*, **MTT-16**, 625 (1968).

[1042] R. Pettai, *Noise in Receiving Systems*, Wiley, New York, 1984.

[1043] J. R. Lewis, "Factors Involved in Determining the Performance of Digital Satellite Links," *Radio and Electronic Engineer*, **54**, 192 (1984).

[1044] E. Fthenakis, *Manual of Satellite Communications*, McGraw-Hill, New York, 1984.

[1045] M. Richaria, *Satellite Communications Systems*, McGraw-Hill, New York, 1995.

[1046] T. Pratt and C. W. Bostian, *Satellite Communications*, Wiley, New York, 1986.

[1047] L. W. Couch, *Digital and Analog Communication Systems*, 4/e, Macmillan, New York, 1993.

[1048] K. Rohlfs, *Tools of Radio Astronomy*, Springer Verlag, New York, 1986.

[1049] S. C. Bundy, "Noise Figure, Antenna Temperature and Sensitivity Level for Wireless Communication Receivers," *Microwave J.*, March 1998, p.108.

[1050] S. C. Bundy, "Sensitivity Improvements and Associated Benefits of Tower-Top Amplifiers," *Microwave J.*, April 1998, p.88.

Beamwidth, Directivity, and Superdirectivity

[1051] R. S. Elliott, "Beamwidth and Directivity of Large Scanning Arrays," *Microwave J.*, Dec. 1963, p.53, and Jan. 1964, p.74.

[1052] R. J. Stegen, "The Gain-Beamwidth Product of an Antenna," *IEEE Trans. Antennas Propagat.*, **AP-12**, 505 (1964).

[1053] W. W. Hansen and J. R. Woodyard, "A New Principle in Directional Antenna Design," *Proc. IRE*, **26**, 333 (1938).

[1054] R. C. Hansen, "Fundamental Limitations in Antennas," *Proc. IEEE*, **69**, 170 (1981), and "Some New Calculations on Antenna Superdirectivity," *ibid.*, p.1365.

[1055] R. C. Hansen, "Superconducting Antennas," *IEEE Trans. Aerosp. Electr. Syst.*, **AES-26**, 345 (1990).

[1056] R. P. Haviland, "Supergain Antennas: Possibilities and Problems," *IEEE Antennas and Propagation Mag.*, **37**, no.4, 13, August 1995.

[1057] R. L. Pritchard, "Maximum Directivity Index of a Linear Point Array," *J. Acoust. Soc. Am.*, **26**, 1034 (1954).

[1058] C. T. Tai, "The Optimum Directivity of Uniformly Spaced Broadside Arrays of Dipoles," *IEEE Trans. Antennas Propagat.*, **AP-12**, 447 (1964).

[1059] Y. T. Lo, S. W. Lee, and Q. H. Lee, "Optimization of Directivity and Signal-to-Noise Ratio of an Arbitrary Antenna Array," *Proc. IEEE*, **54**, 1033 (1966).

[1060] N. Yaru, "A Note on Super-Gain Antenna Arrays," *Proc. IRE*, **39**, 1081 (1951).

[1061] D. R. Rhodes, "The Optimum Line Source for the Best Mean-Square Approximation to a Given Radiation Pattern," *IEEE Trans. Antennas Propagat.*, **AP-11**, 440 (1963).

[1062] D. R. Rhodes, "On an Optimum Line Source for Maximum Directivity," *IEEE Trans. Antennas Propagat.*, **AP-19**, 485 (1971).

[1063] D. K. Cheng, "Optimization Techniques for Antenna Arrays," *Proc. IEEE*, **59**, 1664 (1971).

[1064] W. L. Stutzman, "Estimating Directivity and Gain of Antennas," *IEEE Antennas and Propagation Mag.*, **40**, no.4, 7, August 1998.

Array Design Methods

[1065] H. Bach and J. E. Hansen, "Uniformly Spaced Arrays," in Ref. [9], part 1.

[1066] A. C. Schell and A. Ishimaru, "Antenna Pattern Synthesis," in Ref. [9], part 1.

[1067] S. A. Schelkunoff, "A Mathematical Theory of Linear Arrays," *Bell Syst. Tech. J.*, **22**, 80 (1943).

[1068] C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship Between Beam Width and Side-Lobe Level," *Proc. IRE*, **34**, 335 (1946).

[1069] H. J. Riblet, Discussion of Dolph's paper, *Proc. IRE*, **35**, 489 (1947).

[1070] R. L. Pritchard, "Optimum Directivity Patterns for Linear Point Arrays," *J. Acoust. Soc. Am.*, **25**, 879 (1953).

[1071] R. H. DuHamel, "Optimum Patterns for Endfire Arrays," *Proc. IRE*, **41**, 652 (1953).

[1072] D. Barbiere, "A Method for Calculating the Current Distribution of Tschebyscheff Arrays," *Proc. IRE*, **40**, 78 (1952).

[1073] R. J. Stegen, "Excitation Coefficients and Beamwidths of Tschebyscheff Arrays," *Proc. IRE*, **41**, 1671 (1953).

[1074] C. J. Van der Maas, "A simplified Calculation for Dolph-Tchebyscheff Arrays," *J. Appl. Phys.*, **25**, 121 (1954).

[1075] A. D. Bresler, "A New Algorithm for Calculating the Current Distributions of Dolph-Chebyshev Arrays," *IEEE Trans. Antennas Propagat.*, **AP-28**, 951 (1980).

[1076] A. Zielinski, "Matrix Formulation of Dolph-Chebyshev Beamforming," *Pro. IEEE*, **74**, 1799 (1986), and *ibid.*, **77**, 934 (1989).

[1077] P. Simon, Private Communication, 2003. I would like thank Dr. Simon for permitting me to include his function `chebarray` in this book's MATLAB toolbox.

[1078] C. J. Drane, "Dolph-Chebyshev Excitation Coefficient Approximation," *IEEE Trans. Antennas Propagat.*, **AP-12**, 781 (1964).

[1079] H. D. Helms, "Nonrecursive Digital Filters: Design Methods for Achieving Specifications on Frequency Response," *IEEE Trans. Audio Electroacoust.*, **AU-16**, 336 (1968).

[1080] H. D. Helms, "Digital Filters with Equiripple or Minimax Response," *IEEE Trans. Audio Electroacoust.*, **AU-19**, 87 (1971).

[1081] M. A. Burns, S. R. Laxpati, and J. P. Shelton, Jr., "A Comparative Study of Linear Array Synthesis Using a Personal Computer," *IEEE Trans. Antennas Propagat.*, **AP-32**, 884 (1984).

[1082] F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, **66**, 51 (1978).

[1083] N. C. Ceckinli and D. Yavuz, "Some Novel Windows and a Concise Tutorial Comparison of Window FAmilies," *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-26**, 501 (1978).

[1084] A. H. Nuttal, "Some Windows with Very Good Sidelobe Behavior," *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-29**, 84 (1981).

[1085] T. T. Taylor, "One Parameter Family of Line Sources Producing $\sin\pi u/\pi u$ Patterns," Technical Memorandum no.324, Hughes Aircraft Company, Sept. 1953.

[1086] T. T. Taylor, "Design of Line-Source Antennas for Narrow Beamwidth and Low Side Lobes," *IRE Trans. Antennas Propagat.*, **AP-3**, 16 (1955).

[1087] R. W. Bickmore and R. J. Spellmire, "A Two-Parameter Family of Line Sources," Technical Memorandum no.595, Hughes Aircraft Company, Oct. 1956.

[1088] R. S. Elliott, "Beamwidth and Directivity of Large Scanning Arrays," *Microwave J.*, **6**, 53, December 1963; and *ibid.*, **7**, 74, January 1964.

[1089] J. F. Kaiser, "Nonrecursive Digital Filter Design Using the I_0-Sinh Window Function," *Proc. 1974 IEEE Int. Symp. on Circuits and Systems*, p.20, (1974), and reprinted in *Selected Papers in Digital Signal Processing, II*, edited by the Digital Signal Processing Committee and IEEE ASSP, IEEE Press, New York, 1976, p.123.

[1090] J. F. Kaiser and R. W. Schafer, "On the Use of the I_0-Sinh Window for Spectrum Analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-28**, 105 (1980).

[1091] D. R. Rhodes, "The Optimum Line Source for the Best Mean-Square Approximation to a Given Radiation Pattern," *IEEE Trans. Antennas Propagat.*, **AP-11**, 440 (1963).

[1092] A. Papoulis and M. S. Bertram, "Digital Filtering and Prolate Functions," *IEEE Trans. Circuit Th.*, **CT-19**, 674 (1972).

[1093] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—V: The Discrete Case," *Bell Syst. Tech. J.*, **57**, 1371 (1978).

[1094] S. Prasad, "On an Index for Array Optimization and the Discrete Prolate Spheroidal Functions," *IEEE Trans. Antennas Propagat.*, **AP-30**, 1021 (1982).

[1095] D. Slepian, "Some Comments on Fourier Analysis, Uncertainty and Modeling," *SIAM Rev.*, **25**, 379 (1983).

[1096] J. D. Mathews, J. K. Breakall, and G. K. Karawas, "The Discrete Prolate Spheroidal Filter as a Digital Signal Processing Tool," *IEEE Trans. Acoust., Speech, Signal Process.*, **ASSP-33**, 1471 (1985).

[1097] A. T. Walden, "Accurate Approximation of a 0th Order Discrete Prolate Spheroidal Sequence for Filtering and Data Tapering," *Sig. Process.*, **18**, 341 (1989).

[1098] J. W. Adams, "A New Optimal Window," *IEEE Trans. Acoust., Speech, Signal Process.*, **39**, 1753 (1991).

[1099] J. M. Varah, "The Prolate Matrix," *Lin. Alg. Appl.*, **187**, 269 (1993).

[1100] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge Univ. Press., Cambridge, 1993.

[1101] T. Verma, S. Bilbao, and T. H. Y. Meng, "The Digital Prolate Spheroidal Window," *IEEE Int. Conf. Acoust., Speech, Signal Process.*, **ICASSP-96**, 1351 (1996).

[1102] A. T. Villeneuve, "Taylor Patterns for Discrete Arrays," *IEEE Trans. Antennas Propagat.*, **AP-32**, 1089 (1984).

[1103] J. Butler,"Multiple Beam Antennas," Internal Memo RF-3849, Jan. 1960, Sanders Associates, Nashua, N. H.

[1104] J. P. Shelton and K. S. Kelleher, "Multiple Beams from Linear Arrays," *IEEE Trans. Antennas Propagat.*, **AP-9**, 154 (1961).

[1105] J. L. Allen, "A Theoretical Limitation on the Formation of Lossless Multiple Beams in Linear Arrays," *IEEE Trans. Antennas Propagat.*, **AP-9**, 350 (1961).

[1106] H. J. Moody, "The Systematic Design of the Butler Matrix," *IEEE Trans. Antennas Propagat.*, **AP-12**, 786 (1964).

[1107] J. P. Shelton, "Fast Fourier Transform and Butler Matrices," *Proc. IEEE*, **56**, 350 (1968).

[1108] W. H. Nester, "The Fast Fourier Transform and the Butler Matrix," *IEEE Trans. Antennas Propagat.*, **AP-16**, 360 (1968).

[1109] R. J. Mailloux and H. L. Southall, "The Analogy Between the Butler Matrix and the Neural-Network Direction-Finding Array," *IEEE Antennas Propagat. Magazine*, **39**, no.6, 27 (1997).

Diffraction Theory and Apertures

[1110] Lord Rayleigh, "On the Passage of Waves through Apertures in Plane Screens and Allied Topics," *Phil. Mag.*, **43**, 259 (1897), and *Theory of Sound*, vol. 2, Dover Publ., New York, 1945.

[1111] A. Sommerfeld, *Optics*, Academic Press, New York, 1954.

[1112] S. A. Schelkunoff, "Some Equivalent Theorems of Electromagnetics and their Application to radiation Problems," *Bell Sys. Tech. J.*, **15**, 92 (1936).

[1113] J. A. Stratton and L. J. Chu, "Diffraction Theory of Electromagnetic Waves," *Phys. Rev.*, **56**, 99 (1939).

[1114] S. A. Schelkunoff, “On Diffraction and Radiation of Electromagnetic Waves,” *Phys. rev.*, **56**, 308 (1939).

[1115] K-M. Chen, “A Mathematical Formulation of the Equivalence Principle,” *IEEE Trans. Antennas Propagat.*, **AP-37**, 1576 (1989).

[1116] J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941.

[1117] S. Silver, “Microwave Aperture Antennas and Diffraction Theory,” *J. Opt. Soc. Am.*, **52**, 131 (1962).

[1118] S. Silver, “Radiation from Current Distributions,” “Wavefronts and Rays,” “Scattering and Diffraction,” in Ref. [21].

[1119] F. Kottler, “Electromagnetische Theorie der Beugung an schwarzen Schirmen,” *Ann. der Physik*, **71**, 457 (1923).

[1120] F. Kottler, “Diffraction at a Black Screen,” in E, Wolf, ed., *Progress in Optics*, vol. VI, North-Holland Publishing Co., Amsterdam, 1971.

[1121] W. Franz, “Zur Formulierung des Huygensschen Prinzips,” *Zeit. Naturforschung*, **3a**, 500 (1948).

[1122] J. S. Asvestas, “Diffraction by a Black Screen,” *J. Opt. Soc. Am.*, **65**, 155 (1975).

[1123] C. T. Tai, “Kirchhoff Theory: Scalar, Vector, or Dyadic?,” *IEEE Trans. Antennas Propagat.*, **20**, 114 (1972).

[1124] A. Ishimaru, *Electromagnetic Wave Propagation, Radiation, and Scattering*, Prentice Hall, Upper Saddle River, NJ, 1991.

[1125] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens’ Principle*, 2nd ed., Clarendon Press, Oxford, 1950.

[1126] C. J. Bouwkamp, “Diffraction Theory,” *Repts. Progr. Phys.*, **17**, 35 (1954).

[1127] H. G. Kraus, “Huygens-Fresnel-Kirchhoff Wave-Front Diffraction Formulation,” *J. Opt. Soc. Am.*, **A-6**, 1196 (1989), and **A-7**, 47 (1990), and **A-9**, 1132 (1992).

[1128] S. Ganci, “An Experiment on the Physical reality of Edge-Diffracted Waves,” *Am. J. Phys.*, **57**, 370 (1989).

[1129] T. W. Mayes and B. F. Melton, “Fraunhofer Diffraction of Visible Light by a Narrow Slit,” *Am. J. Phys.*, **62**, 397 (1994).

[1130] J. Durnin, “Exact Solutions for Nondiffracting Beams. I. The Scalar Theory,” *J. Opt. Soc. Am.*, **A-4**, 651 (1987).

[1131] J. Durnin, “Diffraction-Free Beams,” *Phys. rev. Lett.*, **58**, 1499 (1987).

[1132] J. Boersma, “Computation of Fresnel Integrals,” *Math. Comp.*, **14**, 380 (1960).

[1133] ITU Recommendation, “Propagation by Diffraction,” **ITU-R P.526-5**, (1997). Available from [1315].

Geometrical Theory of Diffraction

[1134] J. B. Keller, “Diffraction by an Aperture,” *J. Appl. Phys.*, **28**, 426 (1957).

[1135] J. B. Keller, R. M> Lewis, and B. D. Seckler, “Diffraction by an Aperture. II,” *J. Appl. Phys.*, **28**, 570 (1957).

[1136] J. B. Keller, “Geometrical Theory of Diffraction,” *J. Opt. Soc. Am.*, **52**, 116 (1962).

[1137] R. G. Kouyoumjian and P. H. Pathak, “A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface,” *Proc. IEEE*, **62**, 1448 (1974).

[1138] R. G. Kouyoumjian, “The Geometrical Theory of Diffraction and Its Application,” in Ref. [1224].

[1139] G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, P. Peregrinus, Ltd., England, 1976.

[1140] V. A. Borovikov and B. Y. Kinber, *Geometrical Theory of Diffraction*, IEE Press, London, 1994.

[1141] J. Deygout, “Multiple Knife-Edge Diffraction of Microwaves,”, *IEEE Trans. Antennas Propagat.*, **AP-14**, 480 (1966).

[1142] K, Furutsu, “A Systematic Theory of Wave Propagation Over Irregular Terrain,” *Radio Sci.*, **17**, 1037 (1982).

[1143] L. E. Vogler, “An Attenuation Function for Multiple Knife-Edge Diffraction,” *Radio Sci.*, **17**, 1541 (1982).

[1144] R. J. Luebbers, “Finite Conductivity Uniform GTD Versus Knife-Edge Diffraction in Prediction of Propagation Path Loss,” *IEEE Trans. Antennas Propagat.*, **AP-32**, 70 (1984).

[1145] C. L. Giovaneli, “An Analysis of Simplified Solutions for Multiple Knife-Edge Diffraction,” *IEEE Trans. Antennas Propagat.*, **AP-32**, 297 (1984).

[1146] J. Walfisch and H. L. Bertoni, “A Theoretical Model of UHF Propagation in Urban Environments,” *IEEE Trans. Antennas Propagat.*, **AP-36**, 1788 (1988).

[1147] L. R. Maciel, H. L. Bertoni, and H. H. Xia, “Unified Approach to Prediction of Propagation Over Buildings for All Ranges of Base Station Antenna Height,” *IEEE Trans. Vehic. Tech.*, **VT-42**, 41 (1993).

Plane-Wave Spectrum and Fourier Optics

[1148] D. S. Jones, *The Theory of Electromagnetism*, Macmillan, New York, 1964.

[1149] P. Clemmow, *The Plane Wave Spectrum Representation of Electromagnetic Fields*, Pergamon Press, Oxford, 1966.

[1150] A. Papoulis, *Systems and Trasnforms with Applications in Optics*, McGraw-Hill, New York, 1968.

[1151] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, Wiley, New York, 1991.

[1152] C. Scott, *Introduction to Optics and Optical Imaging*, IEEE Press, New York, 1998.

[1153] T. B. Hansen and A. D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields*, IEEE Press, New York, 1999.

[1154] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform*, Wiley, New York, 2001.

[1155] M. Mansuripur, *Classical Optics and Its Applications*, Cambridge Univ. Press, Cambridge, UK, 2002.

[1156] J. W. Goodman, *Introduction to Fourier Optics*,3d ed., Roberts & Co., Englewood, CO, 2005.

[1157] H. G. Booker and P. C. Clemmow, “The Concept of an Angular Spectrum of Plane Waves and its Relation to that of Polar Diagram and Aperture Distribution,” *Proc. IEEE*, **97**, 11 (1950).

[1158] A. Papoulis, “Ambiguity Function in Fourier Optics,” *J. Opt. Soc. Am.*, **64**, 779 (1974).

[1159] P. P. Banerjee, “On a Simple Derivation of the Fresnel Diffraction Formula and a Transfer Function Approach to Wave Propagation,” *Am. J. Phys.*, **58**, 576 (1990).

[1160] A. Papoulis, “Pulse Compression, Fiber Communications, and Diffraction: A Unified Approach,” *J. Opt. Soc. Am.*, **A-11**, 3 (1994).

[1161] M. Mansuripur, “Fourier Optics, Part 1,” *Opt. & Photon. News*, **11**, p.53, May 2000, and “Fourier Optics, Part 2,” *ibid.*, p.44, June 2000.

Reflector Antennas and Feeds

[1162] L. J. Chu, “Calculation of the Radiation Properties of Hollow Pipes and Horns,” *J. Appl. Phys.*, **11**, 603 (1940).

[1163] J. R. Risser, “Waveguide and Horn Feeds,”, in Ref. [21].

[1164] A. W. Love, ed., *Electromagnetic Horn Antennas*, IEEE Press, New York, 1976.

[1165] P. J. B. Clarricoats and A. D. Olver, *Corrugated Horns for Microwave Antennas*, IEE Electromagnetic Waves Series 18, P. Peregrinus, Ltd., London, 1984.

[1166] S. Silver, “Aperture Illumination and Antenna Patterns,” “Pencil-Beam and Simple Fanned-Beam Antennas,” in Ref. [21].

[1167] R. E. Collin, “Radiation from Apertures,” in Ref. [9], part 1.

[1168] A. D. Yaghjian, “Equivalence of Surface Current and Aperture Field Integrations for Reflector Antennas,” *IEEE Trans. Antennas Propagat.*, **AP-23**, 1355 (1984).

[1169] W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas*, Academic Press, New York, 1970.

[1170] A. W. Love, ed., *Reflector Antennas*, IEEE Press, New York, 1978.

[1171] P. J. Wood, *Reflector Antenna Analysis and Design*, Peter Peregrinus Ltd., London, 1986.

[1172] P. J. B. Clarricoats and G. T. Poulton, “High-Efficiency Microwave Reflector Antennas—A Review,” *Proc. IEEE*, **65**, 1470 (1977).

[1173] W. V. T. Rusch, "Current State of the Reflector Antenna Art—Entering the 1990's," *Proc. IEEE*, **80**, 113 (1992).

[1174] R. C. Johnson, *Designer Notes for Microwave Antennas*, Artech House, Norwood, MA, 1991.

[1175] M. K. Komen, "Use Simple Equations to Calculate Beamwidth," *Microwaves*, Dec. 1981, p. 61.

[1176] A. W. Love, "Some Highlights in Reflector Antenna Development," *Radio Sci.*, **11**, 671 (1976). Reprinted in Ref. [1170].

[1177] C. C. Cutler, "Parabolic-Antenna Design for Microwaves," *Proc. IRE*, **35**, 1284 (1947).

[1178] E. M. T. Jones, "Paraboloid Reflector and Hyperboloid Lens Antennas," *IEEE Trans. Antennas Propagat.*, **AP-2**, 119 (1954).

[1179] J. F. Kauffman, W. F. Croswell, and L. J. Jowers, "Analysis of the Radiation Patterns of Reflector Antennas," *IEEE Trans. Antennas Propagat.*, **AP-24**, 53 (1976).

[1180] A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Trans. Antennas Propagat.*, **AP-21**, 116 (1973). Reprinted in Ref. [1170].

[1181] P. W. Hannan, "Microwave Antennas Derived from the Cassegrain Telescope," *IEEE Trans. Antennas Propagat.*, **AP-9**, 136 (1961). Reprinted in Ref. [1170]

Microstrip Antennas

[1182] K. R. Carver and J. W. Mink, "Microstrip Antenna Technology," *IEEE Trans. Antennas Propagat.*, **AP-29**, 2 (1981).

[1183] A. G. Derneryd, "Linearly Polarized Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, **AP-24**, 846 (1976).

[1184] P. Hammer, D. Van Bouchaute, D. Verschraeven, and A. Van de Capelle, "A Model for Calculating the Radiation Field of Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, **AP-27**, 267 (1979).

[1185] Y. T. Lo, D. Solomon, and W. F. Richards, "Theory and Experiment on Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, **AP-27**, 137 (1979).

[1186] S. L. Chuang, L. Tsang, J. A. Kong, and W. C. Chew, "The Equivalence of the Electric and Magnetic Surface Current Approaches in Microstrip Antenna Studies," *IEEE Trans. Antennas Propagat.*, **AP-28**, 1980.

[1187] I. J. Bahl and P. Bhartia, *Microstrip Antennas*, Artech House, Dedham, MA 1980.

[1188] D. M. Pozar, "Microstrip Antennas," *Proc. IEEE*, **80**, 79 (1992).

[1189] J. F. Zurcher and F. Gardiol, *Broadband Patch Antennas*, Artech House, Dedham, MA, 1995.

Propagation Effects

[1190] S. R. Saunders, *Antennas and Propagation for Wireless Communication Systems*, Wiley, Chichester, England, 1999.

[1191] J. R. Wait, "The Ancient and Modern History of EM Ground-Wave Propagation,", *IEEE Antennas and Propagation Mag.*, **40**, no.5, 7, Oct. 1998.

[1192] J. D. Parsons, *The Mobile Radio Propagation Channel*, Halsted Press, Wiley, New York, 1991.

[1193] T. Maclean and Z. Wu, *Radiowave Propagation Over Ground*, Chapman & Hall, London, 1993.

[1194] K. Siwiak, *Radiowave Propagation and Antennas for Personal Communications*, 2nd ed., Artech House, Norwood, MA, 1998.

[1195] G. Collins, "Wireless Wave Propagation," *Microwave J.*, July 1998, p.78.

[1196] H. T. Friis, A. B. Crawford, and D. C. Hogg, "A Reflection Theory for Propagation Beyond the Horizon," *Bell Syst. Tech. J.*, **36**, 627 (1957).

[1197] J. V. Evans, "Satellite Systems for Personal Communications," *IEEE Antennas Propagat. Magazine*, **39**, no.3, 7, (1997).

[1198] G. Feldhake, "Estimating the Attenuation Due to Combined Atmospheric Effects on Modern Earth-Space Paths," *IEEE Antennas Propagat. Magazine*, **39**, no.4, 26, (1997).

[1199] T. S. Rappaport and S. Sandhu, "Radio-Wave Propagation for Emerging Wireless Personal Communication Systems," *IEEE Antennas Propagat. Magazine*, **36**, no.5, 14, (1994).

[1200] R. K. Crane, "Fundamental Limitations Caused by RF Propagation," *Proc. IEEE*, **69**, 196 (1981).

[1201] K. Bullington, "Radio Propagation for Vehicular Communications," *IEEE Trans. Vehic. Tech.*, **VT-26**, 295 (1977).

[1202] K. Bullington, "Radio Propagation Fundamentals," *Bell Syst. Tech. J.*, **36**, 593 (1957).

[1203] K. Bullington, "Radio Propagation at Frequencies Above 30 Megacycles," *Proc. I.R.E.*, **35**, 1122 (1947).

[1204] K. A. Norton, "The Calculation of Ground-Wave Field Intensity Over a Finitely Conducting Spherical Earth," *Proc. I.R.E.*, **29**, 623 (1941).

[1205] K. A. Norton, "The Propagation of Radio Waves Over the Surface of the Earth and in the Upper Atmosphere," part I, *Proc. I.R.E.*, **24**, 1367 (1936), and part II, *ibid*, **25**, 1203 (1937).

[1206] K. A. Norton, "The Physical Reality of Space and Surface Waves in the Radiation Field of Radio Antennas," *Proc. I.R.E.*, **25**, 1192 (1937).

Numerical Methods

[1207] H. C. Pocklington, "Electrical Oscillations in Wires," *Cambridge Phil. Soc. Proc.*, **9**, 324 (1897).

[1208] E. Hallén, "Theoretical Investigations into Transmitting and Receiving Qualities of Antennas," *Nova Acta Regiae Soc. Sci. Upsaliensis*, p.1, January 1938.

[1209] K. K. Mei, "On the Integral Equations of Thin Wire Antennas,", *IEEE Trans. Antennas Propagat.*, **AP-13**, 374 (1965).

[1210] R.W.P. King and T.T. Wu, "Currents, Charges and Near Fields of Cylindrical Antennas," *Radio Sci. J. Res. NBS/USNC-USRI*, **69D**, 429 (1965).

[1211] Y. S. Yeh and K. K. Mei, "Theory of Conical Equiangular-Spiral Antennas Part I–Numerical Technique," *IEEE Trans. Antennas Propagat.*. **AP-15**, 634 (1967).

[1212] J. H. Richmond, "Digital Computer Solutions of the Rigorous Equations for Scattering Problems," *Proc. IEEE*, **53**, 796 (1965).

[1213] R. F. Harrington, "Matrix Methods for Field Problems," *Proc. IEEE*, **55**, 136 (1967).

[1214] R. F. Harrington, *Field Computation by Moment Methods*, Macmillan, New York, 1968.

[1215] L. L. Tsai and C. E. Smith, "Moment Methods in Electromagnetics for Undergraduates," *IEEE Trans. Education*, **E-21**, 14 (1978).

[1216] M. M. Ney, "Method of Moments as Applied to Electromagnetic Problems," *IEEE Trans. Microwave Theory Tech.*, **MTT-33**, 972 (1985).

[1217] E. H. Newman, "Simple Examples of the Method of Moments in Electromagnetics," *IEEE Trans, Educ.*, **31**, 193 (1988).

[1218] W. P. Wheless and L. T. Wurtz, "Introducing Undergraduates to the Moment Method," *IEEE Trans, Educ.*, **38**, 385 (1995).

[1219] R. F. Harrington, "Origin and Development of the Method of Moments for Field Computation," *IEEE Antennas and Propagation Mag.*, **32**, 31, June 1990.

[1220] L. L. Tsai, "A Numerical Solution for the Near and Far Fields of an Annular Ring of Magnetic Current," *IEEE Trans. Antennas Propagat.*, **AP-20**, 569 (1972).

[1221] L. W. Pearson, "A separation of the logarithmic singularity in the exact kernel of the cylindrical antenna integral equation," *IEEE Trans. Antennas Propagat.*, **AP-23**, 256 (1975).

[1222] C. M. Butler, "Evaluation of potential integral at singularity of exact kernel in thin-wire calculations," *IEEE Trans. Antennas Propagat.*, **AP-23**, 293 (1975).

[1223] C. M. Butler and D. R. Wilton, "Analysis of Various Numerical Techniques Applied to Thin-Wire Scatterers," *IEEE Trans. Antennas Propagat.*, **AP-23**, 534 (1975).

[1224] R. Mittra, Ed., *Numerical and Asymptotic Techniques in Electromagnetics*, Springer-Verlag, New York, 1975.

[1225] E. K. Miller and F. J. Deadrick, "Some Computational Aspects of Thin-Wire Modeling," in Ref. [1224].

[1226] D. R. Wilton and C. M. Butler, "Efficient Numerical Techniques for Solving Pocklington's Equation and their Relationships to Other Methods," *IEEE Trans. Antennas Propagat.*, **AP-24**, 83 (1976).

[1227] G. J. Burke and A. J. Poggio, "Numerical Electromagnetics Code (NEC) — Part II: Program Description – Code," Lawrence Livermore Laboratory Report, UCID-18834, January 1981.

[1228] G. A. Thiele, "Wire Antennas," in *Computer Techniques for Electromagnetics*, R. Mittra, Ed., Hemisphere Publishing Corp., Washington, 1987.

[1229] H-M. Shen and T. T. Wu, "The universal current distribution near the end of a tubular antenna," *J. Math. Phys.*, **30**, 2721 (1989).

[1230] R. W. P. King, "Electric Fields and Vector Potentials of Thin Cylindrical Antennas," *IEEE Trans. Antennas Propagat.*, **38**, 1456 (1990).

[1231] D. H. Werner, P. L. Werner, and J. K. Breakall, "Some Computational Aspects of Pocklington's Electric Field Integral Equation for Thin Wires," *IEEE Trans. Antennas Propagat.*, **42**, 561 (1994).

[1232] D. H. Werner, J. A. Huffman, and P. L. Werner, "Techniques for evaluating the uniform current vector potential at the isolated singularity of the cylindrical wire kernel," *IEEE Trans. Antennas Propagat.*, **42**, 1549 (1994).

[1233] W. A. Davies, "Numerical Methods for Wire Structures,", Tech. Report, EE Dept., Virginia Tech, 1995.

[1234] J. A. Crow, "Quadrature of Integrands with a Logarithmic Singularity," *Math. Comp.*, **60**, 297 (1993).

[1235] J. Ma, V. Rokhlin, and S. Wandzura, "Generalized gaussian quadrature rules for systems of arbitrary functions," *Siam J. Numer. Anal.*, **33**. 971 (1996).

[1236] D. H. Werner, "A method of moments approach for the efficient and accurate modeling of moderately thick cylindrical wire antennas," *IEEE Trans. Antennas Propagat.*, **46**, 373 (1998).

[1237] W. X. Wang, "The Exact Kernel for Cylindrical Antenna," *IEEE Trans. Antennas Propagat.*, **39**, 434 (1991).

[1238] S-O. Park and C. A. Balanis, "Efficient Kernel Calculation of Cylindrical Antennas," *IEEE Trans. Antennas Propagat.*, **43**, 1328 (1995).

[1239] R. R. DeLyser, "Using Mathcad in Electromagnetics Education," *IEEE Trans. Education*, **36**, 198 (1996).

[1240] L. F. Canino, et al., "Numerical solution of the Helmholtz equation in 2D and 3D using a high-order Nyström discretization," *J. Comp. Phys.*, **146**, 627 (1998).

[1241] G. Liu and S. Gedney, "High-order Nyström solution of the Volume–EFIE for TE-wave scattering," *Electromagnetics*, **21**, 1 (2001).

[1242] A. F. Peterson, "Application of the Locally Corrected Nyström Method to the EFIE for the Linear Dipole Antenna," *IEEE Trans. Antennas Propagat.*, **52**, 603 (2004).

[1243] A. F. Peterson and M. M. Bibby, "High-Order Numerical Solutions of the MFIE for the Linear Dipole," *IEEE Trans. Antennas Propagat.*, **52**, 2684 (2004).

[1244] S. A. Schelkunoff, *Advanced Antenna Theory*, Wiley, New York, 1952.

[1245] T. T. Wu and R. W. P. King, "Driving Point and Input Admittance of Linear Antennas," *J. Appl. Phys.*, **30**, 74 (1959).

[1246] T. T. Wu, "Introduction to linear antennas," in [9], Part I.

[1247] R. H. Duncan and F. A. Hinchey, "Cylindrical Antenna Theory," *J. Res. NBS, Radio Propagation*, **64D**, 569 (1960).

[1248] D. S. Jones, "Note on the integral equation for a straight wire antenna," *IEE Proc.*, pt. H, **128**, 114 (1981).

[1249] T. K. Sarkar, "A study of various methods for computing electromagnetic field utilizing thin wire integral equations," *Radio Sci.*, **18**, 29 (1983).

[1250] N. Kalyanasundaram, "On the Distribution of Current on a Straight Wire Antenna," *IEE Proc.,* Pt. H, **132**, 407 (1985).

[1251] B. P. Rynne, "The well-posedness of the integral equations for thin wire antennas," *IMA J. Appl. Math.*, **49**, 35 (1992).

[1252] B. P. Rynne, "Convergence of Galerkin Method Solutions of the Integral Equation for Thin Wire Antennas," *Adv. Comput. Math.*, **12**, 251 (2000).

[1253] G. Fikioris and T. T. Wu, "On the Application of Numerical Methods to Hallén's Equation," *IEEE Trans. Antennas Propagat.*, **49**, 383 (2001).

[1254] R. W. P. King, G. J. Fikioris, and R. B. Mack, *Cylindrical Antennas and Arrays*, 2/e, Cambridge University Press, Cambridge, 2002.

[1255] G. Fikioris, "The approximate integral equation for a cylindrical scatterer has no solution," *J. Electromagn. Waves Appl.*, **15**, 1153 (2001).

[1256] G. Fikioris, J. Lionas, and C. G. Lioutas, "The Use of the Frill Generator in Thin-Wire Integral Equations," *IEEE Trans. Antennas Propagat.*, **51**, 1847 (2003).

[1257] P. J. Papakanellos and C. N. Capsalis, "On the combination of the method of auxiliary sources with reaction matching for the analysis of thin cylindrical antennas," *Int. J. Numer. Model.: Electonoc Netw.*, **17**, 433 (2004).

[1258] A. Heldring and J. M. Rius, "Efficient Full-Kernel Evaluation for Thin Wire Antennas," *Microwave Opt. Technol. Lett.*, **44**, 477 (2005).

[1259] F. D. Quesada Pereira, et al., "Analysis of Thick-Wire Antennas Using an Novel and Simple Kernel Treatment," *Microwave Opt. Technol. Lett.*, **46**, 410 (2005)

[1260] D. R. Wilton and N. J. Champagne, "Evaluation and Integration of the Thin Wire Kernel," *IEEE Trans. Antennas Propagat.*, **54**, 1200 (2006).

[1261] A. Mohan and D. S. Weile, "Accurate Modeling of the Cylindrical Wire Kernel," *Microwave Opt. Technol. Lett.*, **48**, 740 (2006).

[1262] M. C. van Beurden and A. G. Tijhuis, "Analysis and Regularization of the Thin-Wire Integral Equation with Reduced Kernel," *IEEE Trans. Antennas Propagat.*, **55**, 120 (2007).

[1263] K. McDonald, "Currents in a Center-Fed Linear Dipole Antenna," available from: `www.hep.princeton.edu/~mcdonald/examples/transmitter.pdf`..

[1264] A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics*, IEEE Press, New York, 1998.

[1265] D. B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering*, Cambridge University Press, Cambridge, 2005.

[1266] A. Bondenson, T. Rylander, and P. Ingelström, *Computational Electromagnetics*, Springer, New York, 2005.

Elliptic Function Computations

[1267] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.

[1268] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 4/e, Academic Press, New York, 1965.

[1269] D. F. Lawden, *Elliptic Functions and Applications*, Springer-Verlag, New York, 1989.

[1270] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists*, Springer-Verlag, New York, 1971.

[1271] H. J. Orchard and A. N. Willson, "Elliptic Functions for Filter Design," *IEEE Trans. Circuits Syst., I*, **44**, 273 (1997).

[1272] S. J. Orfanidis, "High-Order Digital Parametric Equalizer Design", *J. Audio Eng. Soc.*, **53**, 1026 (2005). MATLAB toolbox available from, `http://www.ece.rutgers.edu/~orfanidi/hpeq/`.

[1273] `http://www.ece.rutgers.edu/~orfanidi/ece521/notes.pdf`, contains a short review of elliptic functions.

[1274] `http://www.ece.rutgers.edu/~orfanidi/ece521/jacobi.pdf`, contains excerpts from Jacobis's original treatise, C. G. J. Jacobi, "Fundamenta Nova Theoriae Functionum Ellipticarum," reprinted in *C. G. J. Jacobi's Gesammelte Werke*, vol.1, C. W. Borchardt, ed., Verlag von G. Reimer, Berlin, 1881.

[1275] `http://home.arcor.de/dfcgen/wpapers/elliptic/elliptic.html`, contains a comprehensive discussion of elliptic functions.

Coupled Antennas, Mutual and Self Impedance

[1276] L. Brillouin, "Origin of Radiation Resistance," *Radioélectricité*, **3** 147 (1922).

[1277] A. A. Pistolkors, "Radiation Resistance of Beam Antennae," *Proc. IRE*, **17**, 562 (1929).

[1278] R. Bechmann, "Calculation of Electric and MAgnetic Field Strengths of any Oscillating Straight Conductors," *Proc. IRE*, **19**, 461 (1931).

[1279] R. Bechmann, “On the Calculation of Radiation Resistance of Antennas and Antenna Combinations,” *Proc. IRE*, **19**, 1471 (1931).

[1280] P. S. Carter, “Circuit Relations in Radiating Systems and Applications to Antenna Problems,” *Proc. IRE*, **20**, 1004 (1932).

[1281] A. W. Nagy, “An Experimental Study of Parasitic Wire Reflectors on 2.5 Meters,” *Proc. IRE*, **24**, 233 (1936).

[1282] G. H. Brown, “Directional Antennas,” *Proc. IRE*, **25**, 78 (1937).

[1283] C. T. Tai, “Coupled Antennas,” *Proc. IRE*, **36**, 487 (1948).

[1284] H. E. King, “Mutual Impedance of Unequal Length Antennas in Echelon,” *IEEE Trans. Antennas Propagat.*, **AP-5**, 306 (1957).

[1285] H. C. Baker and A. H. LaGrone, “Digital Computation of the Mutual Impedance Between Thin Dipoles,” *IEEE Trans. Antennas Propagat.*, **AP-10**, 172 (1962).

[1286] J. H. Richmond and N. H. Geary, “Mutual Impedance Between Coplanar-Skew Dipoles,” *IEEE Trans. Antennas Propagat.*, **AP-18**, 414 (1970).

[1287] R. Hansen, “Formulation of Echelon Dipole Mutual Impedance for Computer,” *IEEE Trans. Antennas Propagat.*, **AP-20**, 780 (1972).

[1288] J. H. Richmond and N. H. Geary, “Mutual Impedance of Nonplanar-Skew Sinusoidal Dipoles,” *IEEE Trans. Antennas Propagat.*, **AP-23**, 412 (1975).

[1289] C. W. Chuang, et al., “New Expressions for Mutual Impedance of Nonplanar-Skew Sinusoidal Monopoles,” *IEEE Trans. Antennas Propagat.*, **AP-38**, 275 (1990).

[1290] J. D. Kraus, “The Corner-Reflector Antenna,” *Proc. IRE*, **28**, 513 (1940).

[1291] H. Yagi, “Beam Transmission of Ultra Short Waves,” *Proc. IRE*, **16**, 715 (1928).

[1292] W. Walkinshaw, “Theoretical Treatment of Short Yagi Aerials,” *J. IEE*, **93**, Pt.IIIA, 598 (1946).

[1293] R. M. Fishenden and E. R. Wiblin, “Design of Yagi Aerials,” *Proc. IEE*, **96**, Pt.III, 5 (1949).

[1294] G. A. Thiele, “Analysis of Yagi-Uda-Type Antennas,” *IEEE Trans. Antennas Propagat.*, **AP-17**, 24 (1968).

[1295] D. K. Cheng and C. A. Chen, “Optimum Element Spacings for Yagi-Uda Arrays,” *IEEE Trans. Antennas Propagat.*, **AP-21**, 615 (1973).

[1296] C. A. Chen and D. K. Cheng, “Optimum Element Lengths for Yagi-Uda Arrays,” *IEEE Trans. Antennas Propagat.*, **AP-23**, 8 (1975).

[1297] G. Sato, “A Secret Story About the Yagi Antenna,” *IEEE Antennas and Propagation Mag.*, **33**, 7, June 1991.

[1298] D. H. Preiss, “A Comparison of Methods to Evaluate Potential Integrals,” *IEEE Trans. Antennas Propagat.*, **AP-24**, 223 (1976).

Web Sites

[1299] `www.codata.org`, Committee on Data for Science and Technology (CODATA). Contains most recent values of physical constants, published in Ref. [89].

[1300] `www.ieee.org/organizations/history_center/general_info/lines_menu.html`, R. D. Friedel, *Lines and Waves, An Exhibit by the IEEE History Center*.

[1301] `www.ece.umd.edu/~taylor/frame1.htm`, Gallery of Electromagnetic Personalities.

[1302] `www.ece.umd.edu/~taylor/optics.htm`, L. S. Taylor, Optics Highlights.

[1303] `mintaka.sdsu.edu/GF`, A. T. Young, “An Introduction to Green Flashes,” with extensive bibliography.

[1304] `www.ee.surrey.ac.uk/Personal/D.Jefferies/antennas.html`, Notes on Antennas.

[1305] `www.arrl.org`, American Radio Relay League.

[1306] `www.qth.com/antenna`, The Elmer HAMlet.

[1307] `www.northcountryradio.com`, North Country Radio.

[1308] `www.tapr.org`, Tuscon Amateur Packet Radio.

[1309] `aces.ee.olemiss.edu/`, Applied Computational Electromagnetics Society,

[1310] `emlib.jpl.nasa.gov`, EMLIB Software Library (with many EM links.)

[1311] `www.qsl.net/wb6tpu/swindex.html`, Numerical Electromagnetics Code (NEC) Archives.

[1312] `dutettq.et.tudelft.nl/~koen/Nec/neclinks.html`, NEC links.

[1313] `www-laacg.atdiv.lanl.gov/electromag.html`, Los Alamos Accelerator Code Group (LAACG), Electromagnetic Modeling Software.

[1314] `soli.inav.net/~rlcross/asap/index.html`, ASAP-Antenna Scatterers Analysis Program.

[1315] `www.itu.org`, International Telcommunication Union(formerly CCIR.)

[1316] `www.fcc.gov`, Federal Communications Commission.

[1317] `sss-mag.com/smith.html`, The Smith Chart Page (with links.)

[1318] `www.wirelessdesignonline.com`, Wireless Design Online.

[1319] `www.csdmag.com`, Communication Systems Design Magazine.

[1320] `home.earthlink.net/~jpdowling/pbgbib.html`, J. P. Dowling, H. Everitt, and E. Yablonovitz, Photonic and Acoustic Band-Gap Bibliography.

[1321] `www.sspectra.com/index.html`, Software Spectra, Inc. Contains thin-film design examples.

[1322] `www.therfc.com/attenrat.htm`, Coaxial Cable Attenuation Ratings.

[1323] `www.tm.agilent.com/data/static/eng/tmo/Notes/interactive/`, Agilent, Application Notes Library.

[1324] `www.semiconductor.agilent.com`, Agilent RF & Microwave Products.

[1325] `www.semiconductor.agilent.com`, Agilent Wireless Library.

[1326] `www.sss-mag.com/spara.html`, S-Parameter Archive (with links.)

[1327] `www.ise.pw.edu.pl/~mschmidt/literature/`, M. Schmidt-Szalowski, Literature on Selected Microwave Topics.

[1328] `www.mathworks.com`, MATLAB resources at Mathworks, Inc.

[1329] `www.3m.com/about3M/technologies/lightmgmt`, Giant Birefringent Optics, 3M, Inc.

[1330] `wwwppd.nrl.navy.mil/nrlformulary/nrlformulary.html`, NRL Plasma Formulary.

[1331] `voyager.jpl.nasa.gov`, Voyager web site.

[1332] `deepspace.jpl.nasa.gov/dsn`, NASA’s Deep-Space Network.

[1333] `www.hep.princeton.edu/~mcdonald/examples`, K. McDonald, “Physics Examples and other Pedagogic Diversions”.

Index

3-dB width, 165
9-dB per delta rule, 55

acoustic tube models, 183
admittance matrix, 522
Ampère's law, 1
anisotropic materials, 5
anisotropic media, 129
antennas
  aperture efficiency, 606
  beam efficiency, 617
  beam solid angle, 602
  beamwidth, 599
  bit-error rate, 624
  center-loaded, 859
  circular loop, 652
  communicating, 612
  coupled, 902
  data rate limits, 624
  directivity, 599
  dish, 746
  dish antennas, 606
  dual-reflector, 763
  effective area, 604
  effective length, 610
  EIRP, 600
  equivalent circuits, 608
  free-space loss, 613
  Friis formula, 612
  gain, 599
  gain-beamwidth relationship, 602, 608
  geosynchronous satellite, 626
  ground effects, 776
  half-wave dipole, 642
  Hertzian dipole, 636
  horns, 727
  lens, 766
  linear, 634
  linear wire, 575
  link budget, 628
  load mismatch, 611
  loop, 650
  microstrip, 740
  minimum-scattering, 610
  monopole, 643
  mutual impedance, 905
  noise temperature, 614
  normalized gain, 601
  open-ended waveguides, 723
  parabolic reflectors, 746
  polarization mismatch, 611
  power density, 599
  radar equation, 629
  radiation efficiency factor, 600
  radiation resistance, 639
  radiation vector, 636
  receiving, 598, 859
  reflector, 746
  resonant, 640
  rhombic, 647
  satellite links, 626
  secant antenna gain, 630
  self impedance, 905
  small dipoles, 638
  square loop, 654
  standing-wave, 638
  transmitting, 598
  traveling wave, 645
  vee, 647
  Yagi-Uda, 923
antireflection coatings, 167, 183, 185
  at oblique incidence, 327
aperture antennas, 723
  aperture-field method, 751
  current-distribution method, 751
  directivity of waveguide apertures, 726
  dual-reflector, 763
  horn design, 737
  horn directivity, 734
  horn radiation fields, 729
  horn radiation patterns, 731
  horns, 727
  lens, 766
  microstrip, 740
  open-ended waveguides, 723
  parabolic reflector beamwidth, 748
  parabolic reflector gain, 748
  parabolic reflectors, 746
  reflector radiation patterns, 754
aperture efficiency, 606, 669
aperture-field method for reflector antennas, 751
apertures
  3-dB angles, 672
  aperture efficiency, 669

  circular, 672
  diffraction theory, 675
  directivity of, 668
  effective area of, 668
  extinction theorem, 679
  field equivalence principle, 658
  Fourier optics, 708
  Franz diffraction formulas, 678
  Fresnel diffraction, 682, 708
  geometrical theory of diffraction, 694
  Huygens source, 666
  Huygens-Fresnel principle, 658
  Kirchhoff diffraction formulas, 677
  knife-edge diffraction, 686
  Kottler's formulas, 663, 678
  lenses, 716
  plane wave spectrum, 703
  Poisson's spot, 712
  radiation fields from, 663
  radiation from, 658
  radiation vectors of, 664
  Rayleigh-Sommerfeld theory, 700
  rectangular, 670
  Sommerfeld's solution, 694
  Stratton-Chu diffraction formulas, 677
  uniform, 670
  vector diffraction for, 681
apparent depth, 251
Appleton-Hartree equations, 148
array design methods, 799
  and FIR filters, 804
  binomial, 822
  Blass matrix, 849
  Butler beams, 848
  continuous line sources, 814
  continuous to discrete, 816
  Dolph-Chebyshev, 823
  DSP analogies, 804
  endfire DuHamel, 834
  Fourier series with windowing, 804
  frequency-sampling, 811
  multibeam, 847
  narrow-beam, low-sidelobe, 818
  prolate array, 840
  Riblet method, 828
  sampled current sources, 815
  Schelkunoff's zero-placement, 802
  secant array, 813
  sector beam, 805
  spatial sampling theorem, 815
  Taylor line source, 842
  Taylor n-bar distribution, 842
  Taylor's one-parameter, 836
  Villeneuve, 846
  Woodward-Lawson, 809, 816
  zero mapping, 816
array factor, 770
array pattern multiplication, 770
array processing, 782
array space factor, 770
arrays, 768
  array factor, 770
  array processing, 782
  beamwidth, 794
  beamwidth of uniform array, 788
  coupled two-element, 911
  directivity, 790
  discrete-space Fourier transform, 782
  DSP analogies, 782
  grating lobes, 784
  ground effects, 776
  interferometry, 784
  one-dimensional, 780
  optimum directivity, 791
  parallel dipoles, 914
  parasitic, 920
  pattern multiplication, 770
  prolate matrix, 791
  rectangular window, 787
  sidelobes of uniform array, 789
  space factor, 770
  spatial z-transform, 782
  steering and scanning, 791
  translational phase shift, 768
  uniform, 786
  very large array, 785
  very long base line, 785
  visible region, 782
  Yagi-Uda, 923
atmospheric refraction, 286
attenuation constant, 55
attenuation in waveguides, 365, 377
attenuator noise temperature, 619
available gain circles, 554
available power gain, 538

BAC-CAB rule, 31
backward recursion, *see* layer recursions
backward waves, 38
balanced stubs, 502
bandwidth
  in waveguides, 374
  noise, 614
  of dielectric mirrors, 194
beam efficiency, 617
beam solid angle, 602
beamwidth, 599
biaxial media, 130
bilinear transformation, 214
binomial arrays, 822
birefingent plasmas, 34
birefringence
  circular, 129
  linear, 129
birefringent media, 129
  Brewster and critical angles, 347

  Brewster angle in, 348
  critical angle of incidence in, 347
  giant birefringent optics, 352, 353, 357
  maximum angle of refraction in, 347
  multilayer structures, 352
  reflection and refraction, 343
bite-error rate (BER), 624
Blass matrix, 849
Bloch wavenumber, 192
Boltzmann constant, 614
Bouguer's law, 288
boundary conditions, 7
  for transverse fields, 244
  matching matrix, 244
BPSK modulation, 624
Brewster angle, 256
  for birefringent media, 348
  for lossy media, 258
  in birefringent media, 347
  in negative-index media, 294
broadband matching, 223
Butler beams, 848

cartesian coordinates, 364, 590
causality, 82
  relativistic, 82
cavity resonators, 382
characteristic impedance, 4, 44
charge
  conservation, 10
  density, 1
  relaxation time, 11
charge flux, 9
charge form factor, 590
charge relaxation, 22
Chebyshev transformer, 476
Chebyshev transformers, 223
chiral media, 129, 132
chiral media, dichroic, 137
chirp radar, 110
chirping, 100
circular apertures, 672
circular loop antennas, 652
coaxial cables, 405
collision time, 17
communicating antennas, 612
complementary error function, 624
complex refractive index, 54
complex waves, 63, 258
conductivity model, 20
conductors in magnetic fields, 33
conjugate matching, 473
constitutive relations, 3
  in anisotropic dielectrics, 5
  in chiral media, 132
  in dispersive materials, 5
  in gyroelectric media, 135
  in gyromagnetic media, 135
  in inhomogeneous dielectrics, 5
  in linear isotropic dielectrics, 4
  in nonlinear dielectrics, 5
  in simple dielectrics, 4
  in uniaxial and biaxial media, 130
  in vacuum, 3
continuous line sources, 814
  Taylor's ideal line, 818
  Taylor's one-parameter, 818
coordinate systems, 947
  cartesian, 364, 590
  cylindrical, 364, 590
  spherical, 590
cosine integrals, 949
coupled antennas, 902
coupled mode theory, 464
  fiber Bragg gratings, 466
coupled transmission lines, 453
coupled two-element arrays, 911
critical angle of incidence, 248
  for birefringent media, 347
crosstalk in transmission lines, 459
current-distribution method for reflectors, 751
currents on linear antennas, 852
cutoff wavenumber and frequency, 362
cylindrical coordinates, 364, 590

data rate limits, 624
delta-gap generator, 855
density
  current, 1
  electric flux, 1
  Lorentz force, 3
  magnetic flux, 1
  momentum, 13, 33
  polarization, 8
  surface charge, 7
  surface current, 7
  volume charge, 1
dichroism, linear and circular, 136
dielectric constant, 4
dielectric mirrors, 190
  bandwidth of, 194
  Fabry-Perot resonators, 201
  multiband, 199
  narrow-band transmission filters, 201
  omnidirectional, 196
  shortpass/longpass filters, 200
dielectric model, 17
dielectric polarization, 4
dielectric slab, 160, 304
  half-wave, 163
  quarter-wave, 163
  reflectionless, 163
dielectric waveguides, 384
diffraction integrals, 949
diffraction theory, 675
diffuse reflection and transmission

  Kubelka-Munk model of, 469
dipole moment density, 4
dipole radiation, 655
directive gain, 599
directivity, 599
directivity of apertures, 668
discretization of continuous line sources, 814
dish antennas, 606, 746
dispersion
  anomalous, 19
  intermodal, 6
  material, 6
  normal, 19
  waveguide, 6
dispersion coefficient, 97
dispersion compensation, 101
dispersive materials, 5
displacement current, 1
Dolph-Chebyshev arrays, 823
Dolph-Chebyshev-Riblet arrays, 828
doppler ambiguity, 117
Doppler effect, 66
Doppler radar, 178
Doppler shift, 175, 178
double-stub tuner, 504
Drude model, 21
DSFT, discrete-space Fourier transform, 782
DTFT, discrete-time Fourier transform, 782
dual-reflector antennas, 763
duality transformation, 660
dynamic predictive deconvolution, 183

effective area of an antenna, 604
effective area of apertures, 668
effective length of an antenna, 610
effective noise temperature, 560, 618
efficiency factor, 600
EIRP, effective isotropic radiated power, 600
electric and magnetic dipoles, 577
electric field, 1
electric flux density, 1
electromagnetic frequency bands, 939
elliptic functions, 870
elliptic integrals, 870
endfire DuHamel arrays, 834
energy conservation, 12
energy density, 12, 14, 25, 45
energy flux, 10, 12, 45, 598
energy velocity, 29, 40
equivalent noise temperature, 620
error function, 624
evanescent waves, 249
  penetration depth for, 249
Ewald-Oseen extinction theorem, 582
exponential integrals, 949
extinction theorem, 582, 679

Fabry-Perot interferometer, 201, 306
far-field approximation, 587
Faraday rotation, 129, 136
Faraday's law of induction, 1
Fermat's principle of least time, 279
fiber Bragg gratings, 183, 201, 466
  quarter-wave phase-shifted gratings, 201
fiber, single mode, 98
fiber, standard, 98
field equivalence principle, 658
field intensities, 1
fields of dipoles, 577
fields of wire antennas, 575
fingerprint identification, 255
flux
  charge, 9
  definition of, 9
  energy, 10, 12
  momentum, 10
  Poynting vector, 12
flux densities, 1
forward recursion, *see* layer recursions
forward waves, 38
Fourier optics, 708
Fourier series method with windowing, 804
Franz diffraction formulas, 678
Fraunhofer region, 589
free-space loss, 613
frequency bands, 939
frequency-sampling array design, 811
Fresnel coefficients, 244, 245
Fresnel diffraction, 682, 708
Fresnel drag, 177
Fresnel integrals, 949
Fresnel region, 589
Fresnel rhomb, 252
Fresnel zones, 689
Friis formula, 612
frill generator, 855
front delay, 82
front velocity, 82
frustrated total internal reflection, 306

gain, 599
gain-beamwidth relationship, 602, 608
Galerkin weighting, 875
gamma-ray bands, 940
gauge transformation, 569
Gauss's laws, 1
geometrical optics, 276
geometrical theory of diffraction, 694
geosynchronous satellite, 603
giant birefringent optics, 137, 352, 353, 357
  mirrors, 353, 354
  reflective polarizers, 356
glass prisms, 250
Goos-Hänchen shift, 253
graded-index optical fibers, 291
grating lobes, 784
gratings, 466
Green's function, 857
Green's function for Helmholz equation, 573
Green's functions, 944
Green's identities, 31
ground effects between antennas, 776
ground-penetrating radar, 78
group delay, 82
group index, 97
group refractive index, 30
group velocity, 29, 30, 80, 82, 92
  in birefringent media, 140
  in waveguides, 375
  slow, fast, and negative, 103
  superluminal, 83, 103
group velocity dispersion, 95
gyroelectric media, 129
gyromagnetic media, 129, 135, 146
gyrotropic media, 135

half-wave dipole antennas, 642
half-wave reflectionless slab, 163
Hall effect, 34, 146
Hallén equation with arbitrary field, 891
Hallén equations for coupled antennas, 929
Hallén integral equation, 854, 857
  approximate kernel, 853
  delta-gap input, 858
  exact kernel, 853
  plane-wave input, 858
harmonic time dependence, 13
Helmholtz equations, 364
Hertzian dipole antenna, 636
high resolution microscopy, 255
horn antennas, 727
horn design, 737
horn directivity, 734
horn radiation fields, 729
horn radiation patterns, 731
Huygens source, 666

impedance
  mutual, 905
  self, 905
  transverse, 63, 240, 242, 362
  wave, 46
impedance matching, 154, 473
  balanced stubs, 502
  Chebyshev transformer, 476
  Chebyshev transformers, 223
  conjugate matching, 473
  double- and triple-stub tuners, 504
  dual-band Chebyshev transformer, 482
  flat line, 474
  L-section matching network, 506
  matching networks, 474
  microstrip matching circuits, 493
  one-section transformer, 498
  Pi-section matching network, 509
  quarter-wavelength transformer, 183, 476
  quarter-wavelength with series section, 488
  quarter-wavelength with shunt stub, 491
  reversed matching networks, 516
  single-stub tuner, 498
  two-section transformer, 493
impedance matrix, 522
impedance transformers, 187
infrared bands, 940
inhomogeneous materials, 5
inhomogeneous waves, 63, 258
integral theorems, 941
intermodal dispersion, 6
internal reflection spectroscopy, 255
inverse power iteration, 842
ionospheric refraction, 283
isotropic radiator, 599

Kaiser window, 807, 836
kernel
  approximate, 853
  elliptic function representation, 870
  exact, 853
  numerical evaluation, 869
  reduced, 853
King's four-term approximation, 868
King's three-term approximation, 862
Kirchhoff diffraction formulas, 677
knife-edge diffraction, 686
Kottler's formulas, 663
Kramers-Kronig dispersion relations, 6, 26, 34
Kubelka-Munk model, 469

L-section matching network, 506
layer recursions, 213, 301
  backward, 214
  forward, 210, 301
left-handed media, 7, 70, 292
lens antennas, 766
lenses, 716
Levinson recursion, *see* layer recursions
linear antennas, 634
linear prediction, 183
link budget calculation, 628
loop antennas, 650
Lorentz dielectric, 18
Lorentz force, 2
Lorentz transformations, 175, 961
Lorenz gauge condition, 569
loss tangent, 24, 58
lossless bounded real functions, 183
lossy media, 52
lossy media, weakly, 58
lossy multilayer structures, 302
low-noise, high-gain, amplifier, 622
lowest usable frequency (LUF), 284

Macneille polarizers, *see* reflective polarizers
magnetic currents, 658, 660
magnetic field, 1
magnetic flux density, 1
magnetic induction, 1
magnetic resonance, 136
magnetization, 4
magnetization current, 6
matched filter, 119
matching, *see* impedance matching
matching matrices, 154, 184
matching matrix, 244
matching networks, 474
material dispersion, 6
MATLAB functions:
  `Cin`, cosine integral Cin, 911, 954
  `Ci`, cosine integral Ci, 911, 954
  `Gi`, Green's function integral, 909, 954
  `RLCmovie`, reactive termination, 442
  `Si`, sine integral, 911, 954
  `TDRmovie`, time-domain reflectometry, 452
  `abp2`, polar gain in absolute units, 919
  `abp`, polar gain in absolute units, 602
  `abz2`, azimuthal gain in absolute units, 919
  `abz`, azimuthal gain in absolute units, 602
  `asinhc`, inverse hyperbolic sinh, 838
  `binomial`, binomial array, 823
  `bkwrec`, backward layer recursion, 219
  `blockmat`, manipulate block matrices, 933
  `brewster`, Brewster and critical angles, 348
  `bwidth`, array beamwidth, 795
  `c2p`, cartesian to phasor form, 562
  `chebarray`, Dolph-Chebyshev array, 827
  `chebtr2`, Chebyshev transformer, 227, 477
  `chebtr3`, Chebyshev transformer, 227, 477
  `chebtr`, Chebyshev transformer, 227, 477
  `dbp2`, polar gain in dB, 919
  `dbp`, polar gain in dB, 602
  `dbz2`, azimuthal gain in dB, 775
  `dbz`, azimuthal gain in dB, 602
  `dguide`, TE modes in dielectric slab, 391
  `diffint`, diffraction integrals, 730, 952
  `dipdir`, dipole directivity, 641
  `dipmovie`, radiating dipole movie, 582
  `dipole`, dipole gain, 643
  `dmax`, dipole directivity, 643
  `dnv`, elliptic function dn, 871
  `dolph2`, Dolph-Chebyshev array, 831
  `dolph3`, Dolph-Chebyshev array, 835
  `dolph`, Dolph-Chebyshev array, 827
  `dsinc`, double sinc function, 725
  `dslab`, cutoff wavenumbers in slab, 391
  `dualband`, dual-band transformer, 486
  `dualbw`, dual-band bandwidth, 488
  `ellipE`, elliptic integral of 2nd kind, 871
  `ellipK`, elliptic integral of 1st kind, 871
  `ellipse`, polarization ellipse, 52
  `fcs2`, Fresnel integrals, 950
  `fcs`, Fresnel integrals, 950
  `fresnel`, Fresnel coefficients, 247, 347
  `frwrec`, forward layer recursion, 219
  `gain1d`, one-dimensional array gain, 771
  `gain2d`, gain of 2D array of dipoles, 934
  `gain2s`, gain of sinusoidal dipole array, 917
  `gin`, input reflection coefficients, 534
  `gout`, output reflection coefficients, 534
  `gprop`, propagation of Γ, 416
  `grvmovie1`, pulse propagation with vg<0, 110
  `grvmovie2`, pulse propagation with vg>c, 110
  `hband`, horn bandedges, 730
  `hbasis`, Hallén basis functions, 891
  `hcoupled2`, coupled Hallén equations, 933
  `hcoupled`, coupled Hallén equations, 933
  `hdelta`, Hallén equation with delta-gap, 880
  `heff`, horn aperture efficiency, 730
  `hfield`, Hallén with arbitrary field, 893
  `hgain`, horn gain patterns, 730
  `hmat`, Hallén impedance matrix, 880
  `hopt`, optimum horn design, 730
  `hsigma`, horn σ parameter, 730
  `hwrap`, wrapped impedance matrix, 880
  `impedmat`, mutual impedance matrix, 917
  `imped`, dipole impedance, 639
  `k2k`, converts to King's primed form, 863
  `kernel`, exact and approximate kernel, 871
  `kingeval`, King's three-term evaluation, 867
  `kingfit`, King's three-term fit, 867
  `king`, Kings three-term approximation, 865
  `landev`, vectorial Landen transformation, 871
  `lmatch`, L-section transformer, 508
  `lmin`, location of voltage min/max, 430
  `mstripa`, microstrip analysis, 404
  `mstripr`, microstrip synthesis, 404
  `mstrips`, microstrip synthesis, 404
  `multbeam`, multibeam array, 848
  `multidiel1`, multilayers, simplified, 302
  `multidiel2`, lossy multilayers, 303
  `multidiel`, multilayer structures, 302, 352
  `multiline`, response of multisection line, 476
  `n2r`, reflection coefficients, 219
  `nfcirc`, noise figure circles, 534
  `nfig`, calculate noise figure, 534
  `omniband2`, birefringent bandwidth, 353
  `omniband`, omnidirectional bandwidth, 333
  `onesect`, one-section transformer, 498
  `p2c`, phasor to cartesian form, 562
  `pfield`, solves Pocklington equation, 898
  `pi2t`, Π to T transformation, 510
  `pmatch`, Π matching network, 512
  `poly2`, improved version of `poly`, 230, 827
  `prolmat`, prolate matrix, 842
  `prol`, prolate array, 842
  `pulse2movie`, pulse on transmission line, 450
  `pulsemovie`, pulse on transmission line, 440
  `quadr2`, Gauss-Legendre quadrature, 961
  `quadrs2`, Gauss-Legendre quadrature, 961

  `quadrs` Gauss-Legendre quadrature, 757
  `quadrs`, Gauss-Legendre quadrature, 955
  `quadr`, Gauss-Legendre quadrature, 955
  `qwt1`, quarter-wavelength transformer, 491
  `qwt2`, quarter-wavelength transformer, 492
  `qwt3`, quarter-wavelength transformer, 492
  `r2n`, refractive indices, 219
  `rhombic`, rhombic antenna gain, 649
  `scan`, array scanning, 793
  `sector`, sector beam design, 809
  `sgain`, calculate power gains, 534
  `sgcirc`, stability and gain circles, 534
  `sinhc`, hyperbolic sinc, 838
  `smatch`, simultaneous conjugate match, 534
  `smithcir`, draw stability or gain circles, 534
  `smith`, draw a basic Smith chart, 534
  `snel`, refraction angle, 346
  `snv`, elliptic function sn, 871
  `sparam`, calculate stability parameters, 534
  `sqrte`, evanescent square root, 261
  `steer`, array steering, 793
  `stub1`, single-stub tuner, 501
  `stub2`, double-stub tuner, 505
  `stub3`, triple-stub tuner, 506
  `swr`, standing wave ratio, 426
  `t2pi`, T to Π transformation, 510
  `taylor1p`, Taylor's one-parameter, 838
  `taylorbw`, Taylor's B-parameter, 838
  `taylornb`, Taylor's n-bar method, 845
  `travel`, traveling-wave antenna, 646
  `tsection`, T-section equivalent, 417
  `twosect`, two-section transformer, 187, 496
  `upulse`, pulse generation, 449
  `ustep`, unit-step generation, 450
  `vee`, vee antenna gain, 649
  `ville`, Villeneuve method, 847
  `vprop`, propagation V, I, 416
  `woodward`, Woodward-Lawson method, 811
  `yagi`, Yagi-Uda array, 925
  `zprop`, propagation of Z, 416
maximum angle of refraction, 248, 347
maximum available gain (MAG), 539
maximum stable gain (MSB), 540
maximum usable frequency (MUF), 284
Maxwell's equations, 1
  harmonic time dependence, 13
  in terms of polarization, 6
  source-free, 2
Maxwell's stress tensor, 33
metamaterials, 7, 292
method of moments, 874
  delta-basis with Galerkin, 882
  delta-function basis, 877
  Galerkin method, 875
  NEC basis, 888
  point matching, 875
  pulse basis, 875, 881
  triangular basis, 886
  weighting functions, 875
microstrip antennas, 740
microstrip matching circuits, 493
microwave amplifier design, 541
microwave frequency bands, 939
microwave oven, 57
mirages, 285
mobility, 22
momentum conservation, 33
momentum density, 13
momentum flux, 10
monopole antennas, 643
moving boundary
  Fresnel drag, 177
  oblique reflection from, 272
  reflection and transmission from, 175
moving media, 175
moving mirror, 178
multibeam array design, 847
multilayer optical film, 357
multilayer structures, 183, 300
  at oblique incidence, 300
  birefringent, 352
  dielectric mirrors, 190, 330
  energy conservation in, 212
  equal-travel time, 206
  lossy, 302
  reflection frequency response of, 211
  scattering matrix of, 212
multiple dielectric slabs, 173
multiple reflections, 171
multisection transmission lines, 475
mutual impedance, 905

narrow-beam, low-sidelobe array design, 818
natural rotation, 134
near fields of linear antennas, 902
negative index media, 7, 30, 70, 292
  Brewster angle in, 294
  equivalent conditions, 71
  propagation in, 70
negative-index media
  lossy, 71
  perfect lens, 319
  Snel's law, 292
network analyzer, 523
noise bandwidth, 614
noise figure, 560, 619
noise figure circles, 560
noise model of a device, 618
noise power, 614
noise temperature, 560, 614
  cellular base station, 615
  of attenuator, 619
  of cascaded devices, 620
  sky, 615
  system, 618
nonlinear materials, 5

normalized gain, 601
numerical aperture, 252
Nyquist frequency
  in multilayer structures, 207

Ohm's law, 6, 21
ohmic power losses, 3, 12, 14, 24
omnidirectional dielectric mirrors, 190, 330
one-dimensional arrays, 780
operating gain circles, 554
operating power gain, 538
optical fibers, 251, 384
  graded index, 291
  numerical aperture, 252
optical filters, 183
  antireflection coatings, 185
  dielectric mirrors, 190
  narrow-band transmission, 201
  shortpass and longpass, 200
optical manhole, 250
optical rotation, 132, 134
optically active media, 129
optimum array directivity, 791

p-polarization, *see* polarization
parabolic reflector antennas, 746
parasitic array, 920
PBG, *see* periodic bandgap structures
penetration depth, 55
  for evanescent waves, 249
perfect electric conductor, 659
perfect lens, 295, 319
perfect magnetic conductor, 659
periodic bandgap structures, 201
  acoustic and vibration control, 201
  Bloch wavenumber, 192
  fiber Bragg gratings, 201
  photonic crystals, 201
  transmission lines and waveguides, 201
permeability, 3
permittivity, 3
phase delay, 82
phase thickness, 207, 301, 330
phase velocity, 82
photonic crystals, 201
physical constants, 938
Pi-section matching network, 509
plane wave incident on linear antenna, 855
plane wave spectrum representation, 703
plasma frequency, 18
plasmas, 25
plasmonics, 272
Pocklington equation solution, 896
Pocklington integral equation, 854
Poisson's spot, 712
polarization, 4, 46, 239
  Brewster angle, 256
  charge density, 8
  linear, circular, 130
  TE, perpendicular, s-polarization, 239
  TM, parallel, p-polarization, 239
polarization current and density, 6
polarization ellipse, 47
polarizers, 131, 341
  beam splitters, 341
  reflective, 341, 356
polarizers, dichroic, 136
polaroid materials, 137
positive real functions, 183
power density, 598
power gain circles, 551
power gains, 536
power losses, 3, 23
power losses in transmission lines, 398
power losses per unit volume, 3
power transfer in transmission lines, 397, 420
power transfer in waveguides, 365
power waves, 542
Poynting vector, 12, 14, 40
precursors, 89
principal-value integrals, 575
prisms, 250
prolate array design, 840
prolate matrix, 791, 840
propagation
  and chirping, 100
  in birefringent media, 129
  in chiral media, 133
  in dispersive media, 80
  in good conductors, 59
  in gyrotropic media, 136
  in negative index media, 70
  in oblique directions, 61
  in uniaxial and biaxial media, 131
  in waveguides, 380
  in weakly lossy media, 58
  matrices, 150
  oblique, in birefringent media, 137
  of reflection coefficient, 151, 244
  of wave impedance, 151, 244
  transient and steady-state, 88
propagation filter, 80
propagation impulse response examples, 85
propagation matrices, 184
propagator frequency response, 81
propagator impulse response, 81
pulse compression, 110
pulse compression filters, 112
pulse compression, and chirping, 101
pulse propagation, 80
pulse propagation, and group velocity, 92
pulse spreading, 6, 95

QPSK modulation, 624
quadruple radiation, 655
quarter-wave reflectionless slab, 163

quarter-wave retarder, 131
quarter-wavelength transformer, 183, 476, 488, 491

radar, 78
radar equation, 629
radar jamming, 633
radar, chirp, 110
radiated power, 599
radiation field approximation, 592
radiation fields, 568, 587, 594
radiation fields from apertures, 663
radiation fields of magnetic currents, 662
radiation from apertures, 658
radiation from dipoles, 577
radiation from waveguide apertures, 723
radiation intensity, 598
  isotropic, 599
radiation patterns of reflector antennas, 754
radiation potentials, 590
radiation resistance, 638, 639
radiation vector, 590
  magnetic, 662
radiation vector of linear antennas, 636
radiative transfer, 469
radio interferometry, 784
radomes, 169
ray tracing, 281
Rayleigh-Sommerfeld diffraction theory, 700
rectangular apertures, 670
recursions, *see* layer recursions
refelection coefficient, 151
reflectance, 163
reflectance spectroscopy, 469
reflected power, 157
reflection by moving boundary, 175
reflection coefficients, 155
  for multilayer structures, 301
  Fresnel, 244, 245
  of multilayer structures, 183
  transverse, 301
reflection response
  forward recursion, 301
  of multilayer structures, 184
reflectionless dielectric slab, 163
reflectionless matching, 473
reflective polarizers, 341
  giant birefringent optics, 356
reflector antennas, 746
refraction
  atmospheric, 286
  ionospheric, 283
  mirages, 285
refractive index, 4, 30, 242
  in birefringent media, 139
  transverse, 242, 301
refractive index model, 582
refractive index, complex, 54
relative permeability, 4
relative permittivity, 4
relativistic causality, 82
relaxation time, 11, 23
relaxation, charge, 22
resonant antennas, 640
resonant cavities, 382
retarded potentials, 570
retarders, 131
reversed matching networks, 516
RF spectrum, 939
rhombic antennas, 647

S-parameters, 522
  available gain, 538
  available gain circles, 554
  generalized, 542
  input and output reflection coefficients, 528
  maximum available gain, 539
  maximum stable gain, 540
  microwave amplifier design, 541
  network analyzers, 523
  noise figure circles, 560
  operating gain, 538
  operating gain circles, 554
  power flow, 526
  power gain circles, 551
  power gains, 536
  power waves, 542
  scattering matrix, 527
  simultaneous conjugate matching, 546
  stability circles, 530
  stability criterion, 533
  transducer gain, 538
  traveling waves, 523
  unilateral gain circles, 552
s-polarization, *see* polarization
satellite links, 626
scattering matrix, 156, 212
  unitarity, 213
scattering parameters, 522
Schelkunoff's zero-placement, 802
Schur algorithm, 183
search radar, 630
sector beam array design, 805
Sellmeier equation, 20
sensors, chemical and biological, 255, 318
Shannon channel capacity, 624
SI units, 1, 938
simultaneous conjugate matching, 546
sine integrals, 949
single-stub tuner, 498
sinusoidal current approximation, 858
skin depth, 55
slef impedance, 905
small dipole antenna, 638
Smith chart, 431
Snel's law, 240, 280

  Bouguer's law, 288
  for lossy media, 246
  in birefringent media, 351
  in multilayer structures, 300
  in negative-index media, 292
solid angle, 592
solitons, 6
Sommerfeld's conducting half-space solution, 694
spatial sampling theorem, 815
spherical coordinates, 590
square loop antennas, 654
stability circles, 530
standard atmosphere, 289
standing wave ratio, 425
standing-wave antennas, 638
stationary phase approximation, 952
Stratton-Chu diffraction formulas, 677
superluminal group velocity, 83, 103
surface current, 60
surface impedance, 60
surface plasmon resonance, 255, 310
surface plasmons, 269, 310
susceptibility, electric, magnetic, 4
system noise temperature, 618
system SNR, 619

Taylor line source array, 842
Taylor one-parameter array design, 836
Taylor's ideal line source, 818
Taylor's one-parameter line source, 818
Taylor-Kaiser arrays, 836
TE and TM impedance, 362
TE waves, 63
TE, TM, TEM modes, 361, 367
telegrapher's equations, 435, 454
thick glasses, 168
thin films, 183, 189, 202
thin-wire kernel, 853
time average
  energy density, 14
  ohmic losses, 14
  Poynting vector, 14
time-domain reflection response, 171
time-domain reflectometry, 451
time-domain response of transmission lines, 435
TM waves, 63
total internal reflection, 247, 261
  critical angle of incidence, 248
  for birefringent media, 347
  frustrated, 306
  maximum angle of refraction, 248
transducer power gain, 538
transfer matrix, 184, 192, 207, 210, 522
transformers, *see* impedance matching
transition matrix, *see* transfer matrix
translational phase shift, 768
transmission coefficients, 155
transmission lines, 394
  broadband terminations of, 183
  cascaded lines, 450
  coaxial lines, 405
  coupled, 453
  coupled telegrapher's equations, 454
  crosstalk, 459
  determination of load impedance, 427
  distributed circuit model of, 412
  equivalent electrostatic problem, 394
  higher modes in, 409
  impedance, inductance, capacitance, 395
  lattice timing diagrams, 438
  microstrip lines, 401
  multisection lines, 475
  open and short circuited lines, 422
  parallel plate lines, 400
  power losses, 398
  power transfer, 420
  reactive terminations, 440
  reflection response, 414
  rise time effects, 449
  Smith chart, 431
  standing wave ratio, 425
  telegrapher's equations, 435
  terminated lines, 417
  Thévenin equivalent circuit, 423, 473
  time-domain reflectometry, 451
  time-domain response, 435
  transient response, 436
  transmitted power, 397
  two-port equivalent circuit of, 416
  two-wire lines, 410
  wave impedance, 414
  weakly coupled lines, 462
transmittance, 163
transmitted power, 157
transverse
  fields, 240
  Fresnel coefficients, 244, 245
  impedance, 242
  propagation matrices, 243
  reflection coefficients, 243, 301
  refractive index, 242, 301
  wave impedance, 243
transverse impedance, 63, 240, 362
traveling wave antennas, 645
traveling waves, 523
triple-stub tuner, 504
two-port network, 522
two-section impedance transformer, 493

ultraviolet bands, 940
uniaxial media, 130
uniform apertures, 670
uniform arrays, 786
uniform plane waves, 36
unilateral gain circles, 552
units, 1

vector diffraction for apertures, 681
vector diffraction theory, 675
vector identities, 941
vector potential, 568
vee antennas, 647
velocity of light, 4
very large array (VLA), 785
very long base line array (VLBA), 785
Villeneuve arrays, 846
visible region, 782
visible spectrum, 940
Voyager spacecraft, 625

wave equation, 37
wave equations, 570
wave impedance, 46, 151, 414
  forward recursion, 302
  of multilayer structures, 184
wavefront, 43, 62
  surfaces and rays, 277
waveguide dispersion, 6
Waveguides
  cutoff wavenumber and frequency, 362
  TE, TM, TEM modes, 361, 367
waveguides, 359
  attenuation, 365, 377
  dielectric, 384
  energy velocity, group velocity, 375
  Helmholtz equations, 364
  operating bandwidth, 374
  power transfer, 365, 375
  propagation model, 380
  rectangular, 370
  resonant cavities, 382
  TE, TM impedance, 362
wavelength division multiplexing, 5
wavelength-division multiplexing, 183, 202
waves
  backward, 38
  complex, 63, 258
  evanescent, 249
  forward, 38
  in birefringent media, 138
  in plasma, 148
  inhomogeneous, 63, 258
  monochromatic, 42
  TE, TM, 63
  uniform plane, 36, 52
  Zenneck, 267
Weyl representation, 945
wire antennas, 634
WMD, *see* wavelength division multiplexing
Woodward-Lawson array design, 809
Woodward-Lawson method, 816

X-ray bands, 940

Yagi-Uda antennas, 923

Zenneck surface wave, 267
zero dispersion wavelength, 98
zero-placement array design, 802